Data and Information Management Plan

Version 1.0
September 2005

National Park Service
Greater Yellowstone Inventory and Monitoring Network

Bighorn Canyon National Recreation Area
Grand Teton National Park
John D. Rockefeller Jr. Memorial Parkway
Yellowstone National Park

NPS GRYN Offices
USGS Northern Rocky Mountain Science Center
229 AJM Johnson
Montana State University
Bozeman, MT 59717

Prepared by

Rob Daley
GRYN Data Manager

With contributions from

Sara Stevens, Margaret Beer, Deborah Angell, Dorothy Mortenson, Chad Jacobson, Ann Rodman, Richard Easterbrook, Cassity Bromley, Shannon Savage, Doug Wilder, Gareth Rowell, Michael Williams, Geoffrey Sanders, Teresa Leibfreid, Bill Moore, Patrick Flaherty, Fred Dieffenbach, Lisa Nelson, Velma Potash, Gary Entsminger, Joe Gregson and Steve Fancy

Web Site (subject to change):
http://www.nature.nps.gov/im/units/gryn/index.shtml

Recommended citation:
Daley R. 2005. Data and Information Management Plan, Greater Yellowstone Inventory and Monitoring Network. Bozeman, MT: National Park Service, Greater Yellowstone Network. 82 p. plus appendices.

Table of Contents

List of Figures

List of Tables

List of Appendices

Executive Summary

Established in 2000 as one of 32 field offices in the NPS Natural Resource Inventory and Monitoring Program, the Greater Yellowstone Network (GRYN) provides coordination and services for the inventory and long term monitoring of selected natural resources in Bighorn Canyon National Recreation Area, Grand Teton National Park (including John D. Rockefeller Jr. Memorial Parkway), and Yellowstone National Park. As a component of the network's Vital Signs Monitoring Plan this data management plan outlines the strategy and guidelines for thorough, integrated, and coordinated resource information management activities to meet NPS-wide information requirements and park level inventory and monitoring information needs. The content of this plan includes overview descriptions, guiding principles, important concepts, and elements of the Service's evolving enterprise architecture for its natural resource line of business. Specific procedures and best practices not covered in this plan are found in the data management and other sections of the network's individual monitoring protocols.

About one-third of I&M program and network resources are devoted to generating, managing, and sharing data and information resources that are well planned, highly organized, easily accessible, completely documented, and scientifically credible. The purpose of emphasizing data management is to provide stable, long-term support for natural resource managers and decision makers in the NPS, as well as non-NPS partners, scientists, and other stakeholders.

The foundation for effective data management begins with specifying data stewardship roles and responsibilities for every aspect of program and project planning and operations. This plan lists several roles and associated responsibilities for data management that cover data producers through end users. Management level support for these responsibilities, particularly at parks, is required to meet I&M program goals. Park managers need to plan for and provide funds and time to adequately manage data and information resources generated by I&M program and other natural resource management activities. Managers at all levels also need to provide accountability for data management responsibilities as performance elements for all natural resource positions. Emphasis is placed on the collaborative nature of successful data stewardship, and the network promotes a teamwork approach to data management. Because it is essential to establish, communicate, and support data stewardship responsibilities for all staff at all stages of a project, chapter three discusses project workflow and data management issues and opportunities. A standard data lifecycle model helps demonstrate the relationship between park-level data collection for inventory and monitoring projects with NPS service-wide databases and data distribution services.

The network operates and maintains a NPS-compliant information technology infrastructure at offices currently located on the Montana State University campus in Bozeman. Hardware, software, security provisions, and local area network services are managed by GRYN staff with support from regional and park level technology service staff. The network maintains project-specific database applications and data sets designed and populated based on inventory study plans and specific monitoring objectives, often in partnership with non-NPS partners such as the US Geological Survey and the US Forest Service. The data stored in these local applications are transferred at appropriate intervals to NPS service-wide

databases and repositories including NPSpecies, NatureBib, NPSTORET, the NPS Natural Resource GIS Metadata and Data Clearinghouse, and the NPS Biodiversity Data Store.

The network necessarily focuses on collecting and managing data that are required to meet monitoring objectives specified in the vital sign monitoring protocols. GRYN also works with staff from network parks, the NPS regional and Washington offices, and other partner entities to identify, manage, and integrate legacy data from past monitoring efforts and relevant data sources from ongoing monitoring that are not based strictly on protocols generated by the NPS I&M program. Chapter five discusses the networks options for recording data in the field, its approach to database design and data discovery, acquisition, and entry, and the treatment of data from different organizational levels in the NPS.

An important data management goal discussed in chapter six is how to ensure that I&M projects produce data that meet quality requirements specified in vital sign monitoring protocols. The plan discusses ways to reduce errors when entering data, procedures for verifying and validating data, and methods to share information about data quality to promote the appropriate application and use of data resources. Another principal component of effective data management operations is data documentation, discussed in the context of network activities involving NPS metadata creation and distribution tools in chapter seven. Chapter eight follows with a brief discussion of the supporting role of data management for analysis and reporting.

GRYN is committed to making all available and relevant data easily accessible while providing for the appropriate level of protection for sensitive data resources. Procedures and solutions for making data available to park natural resource managers and decision makers, as well as other stakeholders and interested parties are presented in Chapter nine. In the final chapter the long term elements of data maintenance, storage, and archiving are addressed for electronic as well as physical (specimen) and hardcopy data and information resources.

Current and future GIS and information management strategies and solutions at each park represent valuable information about what works and what needs improvement in the realm of data management. Staff from the network and the parks will coordinate and integrate data management activities based on common tools, best practices and shared objectives. This plan is a key mechanism to help understand and address the relationships between park level information needs supported by park GIS and data management activities, and I&M program organization and standards for Network data management.

1. Introduction

In 1999 the National Park Service released its Natural Resource Challenge, an Action Plan for Preserving Natural Resources. The Action Plan focuses on ways to expand scientific research to inform the management of park resources, provide benchmarks and repositories of environmental information for local, regional, and global interests, and share knowledge with the scientific community, academic institutions, and the public. One component of the Action Plan involves the NPS Natural Resource Inventory and Monitoring (I&M) Program, which works to address standard natural resource inventories and long term status and trend monitoring for park-selected ecological indicators known as Vital Signs. To coordinate and provide services for the inventory and monitoring activities in four NPS units, the Greater Yellowstone Network (GRYN) was established in 2000 as one of 32 networks of parks in the I&M Program. Bighorn Canyon National Recreation Area, Grand Teton National Park, John D. Rockefeller Jr. Memorial Parkway (administered by Grand Teton NP), and Yellowstone National Park involve nearly 2.6 million acres in parts of Idaho, Montana, and Wyoming (Figure 1.1).

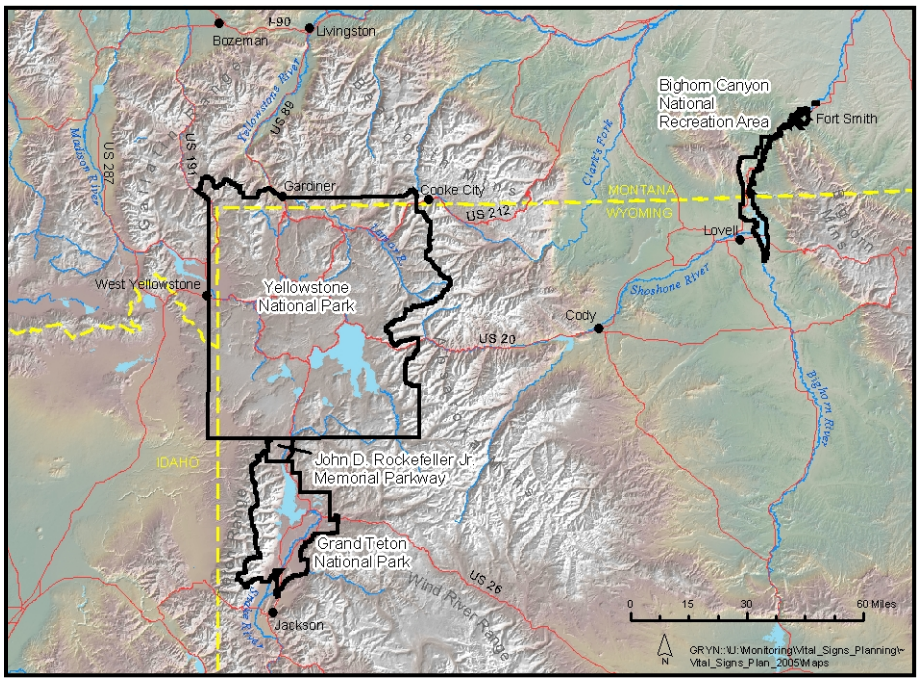


Figure 1.1. National Park Service Units served by the GRYN.

In a paper presented at the 2001 George Wright Society Conference, Abigail Miller of the National Park Service stated the need for a system-wide approach in the NPS to ensure that data meet national level standards for quality and availability. She set expectations for inventory and monitoring networks to strategically approach data collection and management activities that help meet park level information requirements as well as provide credible and useful data and information for landscape level and national efforts

This Data and Information Management Plan (the 'Plan') presents principles and guidelines that reflect the National Park Service Inventory and Monitoring Program's goal to ensure that investments made to acquire scientific data for park natural resources yield high quality data

and information assets that are organized, available, useful, compliant, and secure. To achieve these fundamental requirements, the Plan focuses on the following objectives:

- Provide data management services and guidance in support of the Inventory and Monitoring Program's goal to identify, catalog, organize, structure, archive, and make available relevant natural resource information;
- Initiate and invest in data management activities based on data and information needs defined in network monitoring protocols and inventory study plans;
- Integrate data management activities with all aspects and at all stages of network business;
- Specify data stewardship responsibilities for all personnel;
- Collaborate internally and externally to address data management issues with individuals at all organizational levels.

1.1 GRYN Organization

The network's coordinator, ecologist, and data manager are stationed in Bozeman, Montana. Temporary staff can be stationed in Bozeman and at the parks. Network business is guided locally by a Technical Committee of managers from the network parks, and approved by a Board of Directors made up by superintendents (or their representative) from each network park. Yellowstone National Park and Grand Teton National Park have active GIS and resource information management programs whose staff work with network personnel to identify and develop common strategies and policies for integrated data management. Resource specialists and managers at Bighorn Canyon National Recreation Area work with network staff to develop and implement data management practices that meet the information needs of the recreation area and the network.

1.2 Data and Information Synthesis

Natural resource **data** include collections of estimated, predicted, observed, and measured values or classifications for any element or event of an ecological system, including physical, chemical, biological, and human dimensions. The patterns and relationships among these data represent the **information** that provides meaning to people who wish to understand, manage, and use park resources. An important objective of the network's Vital Signs Monitoring Plan is to synthesize and deliver overall ecosystem status and trend information. Resulting products are based on data from multiple network Vital Sign monitoring protocols and other relevant data and information in various formats and conditions, from sources within and outside the National Park Service. The network manages all types, forms, and stages of data and information and their inherent relationships, provided that they support one or more of the following program elements:

- goals and objectives of the network's Vital Signs Monitoring Plan
- specific needs defined in approved Vital Sign monitoring protocols
- network inventory study plan objectives
- other specific natural resource management projects that network personnel and park staff agree to cooperate in developing and managing

1.3 Goals and Objectives for Managing Natural Resource Information Assets

The overall mission for GRYN data management is to provide and support data and information resources that are organized, useful, available, compliant, and secure. The

network's monitoring efforts and resource information management activities are expected to provide park managers with the scientific data they need to understand and manage park resources. Based on the strategic approach to data management from the Inventory and Monitoring Program, the goals of the network include the following:

- Identify the most critical data needs for parks and networks
- Develop partnerships with other land and resource management agencies, non-government entities, and public and private interests
- Collaborate internally and externally to help establish consistency among approaches and protocols
- Compare and share protocols and data sets locally and via the NPS I&M Protocol Database
- Use standard quantitative and qualitative measures to report on the condition and trends of natural resources

Primary objectives of natural resource data and information management include:

- Manage data and information as programmatic and service-wide resources
- Manage data acquisition to avoid duplication and unnecessary expense
- Store, manage, and maintain data in one or more central and accessible repositories
- Protect data from unauthorized or unplanned modifications
- Perform quality controls on data throughout its life cycle
- Share data and information appropriately with internal and external interests

1.4 Data Management Plan Scope and Revision Schedule

Data and information resources relevant to the network are those that result from or relate to the implementation of a formal network Vital Signs monitoring protocol, network Inventory Study Plan, or park and network approved project with well defined and thoroughly reviewed information needs. This plan covers ways in which the network manages all programmatic information assets, including hardcopy reports and field forms, electronic files, spatial and tabular data sets, image data from remote sensors and cameras, audio and video data, and ground-based images. The plan also discusses the network's involvement with the parks and conservatories in handling and managing physical specimens such as voucher evidence for records used in NPSpecies, the National Park Service's biological diversity database.

This plan, its appendices, and the data management section in each Vital Sign monitoring protocol require updates and regular maintenance by network staff. To provide access to data and information management principles and guidance, the network continuously maintains a web-based data and information management plan. Minor revisions and edits occur regularly and more substantial changes or additions occur as needed when relevant new information results from network business, NPS policy, or Inventory and Monitoring Program direction. A comprehensive review and update of the Data and Information Management Plan will occur every three years. Changes to the Plan are documented in the Revision History Log at the end of this document.

2. Data Stewardship Roles and Responsibilities

The benefits to natural resource managers, scientists, and the public from inventory and monitoring project results are substantially affected by the ability of the individuals working on the project, the network staff, and the National Park Service to keep track of data from the time they are gathered until and while they inform a decision making process. Managing data for multiple monitoring protocols and NPS inventories involves timeframes of several years and includes changes in information technology, turnover in staff, new scientific insights, and shifting priorities. This requires that every individual involved with the Greater Yellowstone network understand and perform data stewardship responsibilities in the production, analysis, management, and/or end use of the data (Table 2.1). As coordinator of these activities, the fundamental role of the network data manager is to understand program and project requirements, create and maintain data management infrastructure and standards, and educate, communicate, and work with all responsible individuals. In order to achieve the highest possible success in delivering useful and credible scientific information about park ecosystems, the Inventory and Monitoring Program expects each network to invest at least thirty percent of available resources (time, money, and expertise) in data management, analysis, and reporting. In addition to the overall data management roles and responsibilities presented here, each Vital Sign monitoring protocol and inventory study plan contains specific instructions for assignments and tasks that nest within a data stewardship framework.

Table 2.1. Categories of data stewardship involving all network personnel.

Stewardship Category	Related Activities	Principal jobs or positions
Note: Each position is listed in only one category according to overriding responsibilities. However, most positions contribute in each category.		
Production	Creating data or information from any original or derived source. This includes recording locations, images, measurements, and observations in the field, digitizing source maps, keying in data from a hardcopy source, converting existing data sources, image processing, and preparing and delivering informative products, such as summary tables, maps, charts, and reports.	Project Crew Member Project Crew Leader Data/GIS Specialist or Technician
Analysis	Using data to predict, qualify, and quantify ecosystem elements, structure, and function as part of the effort to understand these components, address monitoring objectives, and inform park and ecosystem management.	Network Ecologist Resource Specialist Statistician/Biometrician
Management	Preparing and executing policies, procedures, and activities that keep data and information resources organized, available, useful, compliant, and secure.	Network Data Manager Project Leader GIS Manager IT Specialist Database Manager National I&M Data Manager Curator

Stewardship Category	Related Activities	Principal jobs or positions
End Use	Informing the scope and direction of science information needs and activities. Providing the money and scheduling the time required for project leaders and staff to meet stated objectives for data quality and other data management objectives. Obtaining and applying available information to develop knowledge that contributes to understanding and managing park resources. Providing feedback for improvements in data content and quality.	Network Coordinator Park managers and superintendents Others

2.1 Definitions of Role and Responsibility

The network considers a role as a function or position (*e.g., Data Manager*) while a responsibility is a duty or obligation (*e.g., review data records*). Data stewardship is not treated as a role because it involves all project staff sharing the responsibilities for managing data and information resources. Using the term 'Data Steward' seems appropriate only when a list of specific responsibilities is associated with every role in a given project or program. Establishing a single project data steward role is avoided because an unmanageable burden (real or perceived) is placed on a single position/person, and others with data management responsibilities are more likely to under perform.

2.2 Primary Roles and Responsibilities for Data Management

An increasing demand for more detailed, higher quality data and information about natural resources and ecosystem functions requires a group of people working together to effectively steward data and information assets. The current and expected capacity, diversity, and rate of change in information technology make managing any large amount of data a greater task than any person can reasonably do alone. Knowledgeable individuals from scientific, administrative, and technological disciplines must work together to ensure that data are collected using appropriate methods, and that resulting data sets, reports, maps, models, and other derived products are well managed. Data sets and related products must be credible, representative, and available for current and future needs. A typical data set used for inventory and monitoring is handled, viewed, and stewarded by most or all of the people serving in the roles (jobs) listed in Table 2.2. These roles are listed 'from the ground up' to help demonstrate the hierarchy and overlap of responsibilities. For example, a project leader is ultimately responsible for the activities listed in the field level roles of crew leader and crew member, and the network coordinator ensures that the network data manager and ecologist achieve the required performance level. With one third of network resources devoted to overall data management, analysis, and reporting, the network intends to avoid cases in which a single position is assigned unattainable duties for multiple roles, resulting in a lack of attention to one or more aspects of the project. Meeting this goal requires awareness and support from park managers to plan for, fund, and provide performance level accountability for the resources required to adequately manage their park's data and information assets. Learning, understanding, and acting on the responsibilities listed for all categories of data stewardship will help prevent the critical aspects of data management from being overlooked or under-funded.

Table 2.2. Summary of Roles and Responsibilities

Role	Primary responsibilities related to data management
Project Crew Member	Record and verify measurements and observations based on project objectives and protocols. Document methods, procedures and anomalies.
Project Crew Leader	Supervise crew members to ensure their data collection and management obligations are met, including data verification and documentation.
Data/GIS Specialist or Technician	Perform assigned level of technical data management and/or GIS activities, including data entry, data conversion, and documentation. Work on overall data quality and stewardship with project leaders, resource specialists, and the Network data manager.
Information Technology/Systems Specialist	Provide and maintain an information systems and technology foundation to support data management.
Project Leader	Oversee and direct operations, including data management requirements, for one or more network projects. Maintain communication with project staff, Network Data Manager, and resource specialist regarding data management. Determine what others inside and outside the NPS are doing relative to the project, and establish a minimum set of standard data items to promote comparability with other efforts. **Note:** The Project Leader is often a resource specialist, in which case the associated responsibilities for data authority apply (see next role). A Project Leader without the required background to act as an authority for the data will consult with and involve the appropriate resource specialists.
Resource Specialist	Understand the objectives of the project, the resulting data, and their scientific and management relevance. Make decisions about data with regard to validity, utility, sensitivity, and availability. Describe, publish, release, and discuss the data and associated information products. **Note:** The Resource Specialist serving as a Project Leader is also responsible for the duties listed with that role.
GIS Manager	Support park management objectives. Coordinate and integrate local GIS and resource information management with network, regional, and national standards and guidelines.
Network Data Manager	Provide overall network planning, training, and operational support for the awareness, coordination, and integration of data and information management activities, including people, information needs, data, software, and hardware. Serve as Point of Contact for National Park Service database applications (NPSpecies) Coordinate internal and external data management activities, especially in collaboration with project leaders Ensure that metadata is created to meet NPS standards and published on an NSDI clearinghouse
Database Manager	Apply particular knowledge and abilities related to database software and associated application(s)
Curator	Oversee all aspects of the acquisition, documentation, preservation, and use of park collections
Statistician/Biometrician	Analyze data and present information

Role	Primary responsibilities related to data management
Network Ecologist	Ensure useful data are collected and managed by integrating natural resource science in network activities and products, including objective setting, sample design, data analysis, synthesis, and reporting
Network Coordinator	Ensure programmatic data and information management requirements are met as part of overall network business
I&M Data Manager (National Level)	Provide service-wide database design, support, and services, including receiving and processing to convert, store, and archive data in service-wide databases
Other End Users	These 'information consumers' include park managers and superintendents, researchers, staff from other agencies, and the public. End users are responsible for informing the scope and direction of science information needs and activities, applying data and derived products appropriately, and providing feedback for improvements. Park superintendents and program managers, in particular, are responsible for providing the money and scheduling the time required for project leaders and staff to meet stated data quality and other data management objectives.

2.3 Coordination with parks and others

The Natural Resource Challenge states that collaboration among the National Park Service, other public agencies, universities, and non-governmental organizations is necessary to effectively acquire, apply, and promulgate the scientific knowledge gained in parks. The Inventory and Monitoring Program encourages coordination among participants at all levels to help ensure that data collected by NPS staff, cooperators, researchers and others are entered, quality-checked, documented, analyzed, reported, archived, cataloged, and made available for management decision-making, research, and education. The network data manager works with national Inventory and Monitoring Program data management staff and regional resource information management personnel to maintain awareness and involvement in service-wide and regional databases and data management policy and guidance. The network data manager works locally with network personnel, park staff, and cooperators to promote and develop workable standards and procedures that result in integration and availability of data sets. Key contacts for the network data manager include park GIS and data managers and the project leaders for each monitoring or inventory project. Regular and productive communication among these personnel leads to common understanding and better synchronization of network and park data management activities. Park and network staff coordinate on resource information management using a variety of methods, including personal visits, phone calls, email, joint meetings and training sessions, as well as the meetings and work of the network's Technical Committee, Science Committee, and Board of Directors. The development of network planning materials, including inventory study plans and monitoring protocols, includes involvement and input from park scientists and resource information management staff. This data and information management plan, for example, is prepared with input and review from park GIS managers and others.

2.3.1 Network of Networks

The prototype monitoring parks and networks established a foundation of collaboration and coordination in network and park data management that continues to benefit the Inventory and Monitoring Program and data management in the National Park Service.

Data managers throughout the program regularly coordinate with each other and national program staff via annual meetings, conference calls, workgroups, a listserv, web sites, and informal communication. Data managers from several networks divided the workload and collaborated on developing their respective network Data Management Plans. This model of cooperation using available resources and strong communication is effective in the National Park Service and can be productively applied to most situations involving resource information management across and among the administrative units of the organization. The Greater Yellowstone Network maintains an active role in coordination between networks and parks to promote practical consistency among protocols, database designs, and data sets. GIS and data managers from network parks are asked to participate in program development and activities, take advantage of Inventory and Monitoring Program resources, and communicate with network and program staff to share information about progress and direction and address concerns.

2.3.2 External Coordination

The two largest network parks are partners in land and resource management within the Greater Yellowstone Ecosystem, which garners a high level of interest ranging from research and interpretation of the Ecosystem to dialogue about natural resource policy and management issues. This pervasive and intense interest bolsters the network's responsibility and the opportunity to participate in coordinated data management activities with external stakeholders. This involves establishing and maintaining awareness and communication regarding resource information management activities at other federal, state, and local government offices. For example, the Greater Yellowstone Coordinating Committee is an established group of federal land managers who have an interest in promoting high quality data and information management to support the work and products of their issue teams, subcommittees and the associated projects. The network uses and supports the functions of the Greater Yellowstone Coordinating Committee to maintain a role in ecosystem management based on credible data both inside and outside the parks. Another important commitment and opportunity for the network involves interagency collaboration on the development of Vital Sign monitoring protocols, which includes examining and producing solutions to make data useful among different agencies. Most or all network monitoring protocols benefit from coordination with inventory and monitoring interests in the National Park Service and from contributions made by external partners. A list of past and current contacts and contributors is maintained in the network's information system. Partnerships with organizations and agencies outside of the National Park Service are available for reference in the network Vital Signs Monitoring Plan.

2.4 Documentation is Key

If a single shared responsibility stands out in importance and value, it is to document data sets, the data source(s), and the methodology by which the data were collected or acquired. This establishes the basis for the appropriate use of the data in resulting analysis and products, both short term and long term. Network monitoring protocols contain key elements of data documentation. Network data records collected according to these protocols will include the name, date, and version of the associated protocol. Additional important guidance and reference for documentation and metadata are found in the Data Documentation chapter of this plan.

2.5 Project Stewardship

Since the data management aspects of every inventory or monitoring project normally require the expertise and involvement of several people over a period of months or years, it makes sense that one person is charged with keeping track of the objectives, requirements, and progress for each project. This project leader (or steward) is normally a resource management specialist with training and experience in the field of science related to the inventory or monitoring project and may have worked in the geographic area where the study occurs. A background like this prepares the project leader to oversee the field work, coordinate with GIS and data managers on information management needs, understand the project's objectives, and authorize the validity, utility, sensitivity, and availability of data resources. A project leader who lacks the background or expertise to act as an authority for the data will work with the appropriate resource specialists to account for those aspects of data stewardship. In order to ensure quality management of each project, including data management requirements, project leaders are assigned only those projects they can effectively oversee based on workload and other relevant factors determined by the network coordinator. Unless the project is short term (3 months or less) the project will have at least one alternate or backup project leader to provide continuity in case the principal project leader becomes temporarily or completely unavailable. Table 2.2 describes the primary responsibilities of a project leader, resource specialist, and network data manager that, when coordinated for a project, result in quality data and information products. Comprehensive responsibilities and assignments for these roles vary by project and are listed and described in detail with each Vital Sign monitoring protocol or other project study plan.

2.6 Data Stewardship

Keeping track of data from the time of acquisition until it is no longer useful is the shared responsibility of everyone involved with data as a producer, analyst, manager, or end user. This, in essence, is data stewardship. It is a principle of mutual accountability rather than a particular job for one individual. The importance of data stewardship must not be understated, and the degree of success with which it is used will have direct bearing on the quality and utility of products developed by the network. Data stewardship responsibilities for network and park level roles are listed in Appendix A, and are specified with more detail in individual monitoring protocols and study plans.

2.6.1 Awareness and Accountability

Successful data stewardship requires that people involved in network activities learn and understand the expectations for continuous data management, and are held accountable to perform the duties required to meet these expectations. This is equally important for network staff, park employees, and contractors or cooperators. All project participants receive training, briefings, materials, and additional regular communication about data stewardship from supervisors, project leaders, and data managers. The purpose is to promote the appropriate level of understanding about how their efforts relate to park and network management objectives, National Park Service and Department of Interior policies, and other federal government requirements. Other relevant context and linkages can also be discussed to help establish a sense of ownership and accountability among project staff. Inventory and Monitoring project leaders have a good understanding of resource information management issues and requirements and an awareness of the challenges and limitations associated with field data collection, including the use of GPS technology. This is achieved

through training, detailed and regular briefings, and accompanying field crews to perform data collection at reasonable intervals.

2.6.2 The Hub of Data Stewardship

Project leaders, data managers, and GIS specialists comprise the central data management team for inventory and monitoring projects. Each is responsible for certain aspects of project data, and all share responsibility for some overlapping tasks. Because of the collaborative nature of project data management, communication among project leaders, data managers and GIS specialists is essential to meeting program goals. The following section outlines the individual and shared responsibilities of each role.

Project leaders (stewards)

The project leader is accountable for data quality during all phases of the project, including collecting, entering, handling, reviewing, summarizing, and reporting data. Developing project documentation and metadata are crucial elements of the project leader's role. Specifically, the role of the project leader includes the following.

- Complete project documentation describing the who, what, where, when, why and how of a project.
- Develop, document and implement standard procedures for field data collection and data handling.
- Enact and supervise quality assurance and quality control measures for the project. Supervise and certify all field operations, including staff training, equipment calibration, species identification, and data collection. Supervise or perform data entry, verification and validation.
- Maintain concise explanatory documentation of all deviations from standard procedures.
- Ensure documentation of important details for each field data collection period.
- Maintain hard copies of data forms and send original data forms to archive on a regular basis.
- Work with program coordinators to identify analysis and reporting mechanisms and to establish a schedule for regular project milestones such as data collection periods, data processing target dates, and reporting deadlines.
- Produce regular summary reports and conduct periodic trend analysis of data, store the resulting reports, and make them available to users.
- Act as the main point of contact concerning data content.

The project leader will also work closely with the data manager to:

- Develop quality assurance and quality control procedures specific to project operations.
- Identify training needs for staff related to data management philosophy, database software use, quality control procedures, etc.
- Coordinate changes to the field data forms and the user interface for the project database.
- Fully document and maintain master data.
- Identify sensitive information that requires special consideration prior to distribution.
- Manage the archival process to ensure regular archival of project documentation, original field data, databases, reports and summaries, and other products from the project.

- Define how project data will be transformed from raw data into meaningful information and create data summary procedures to automate and standardize this process.
- Identify and prioritize legacy data for conversion; convert priority data sets to a modern format.
- Increase the interpretability and accessibility of existing natural resource information.

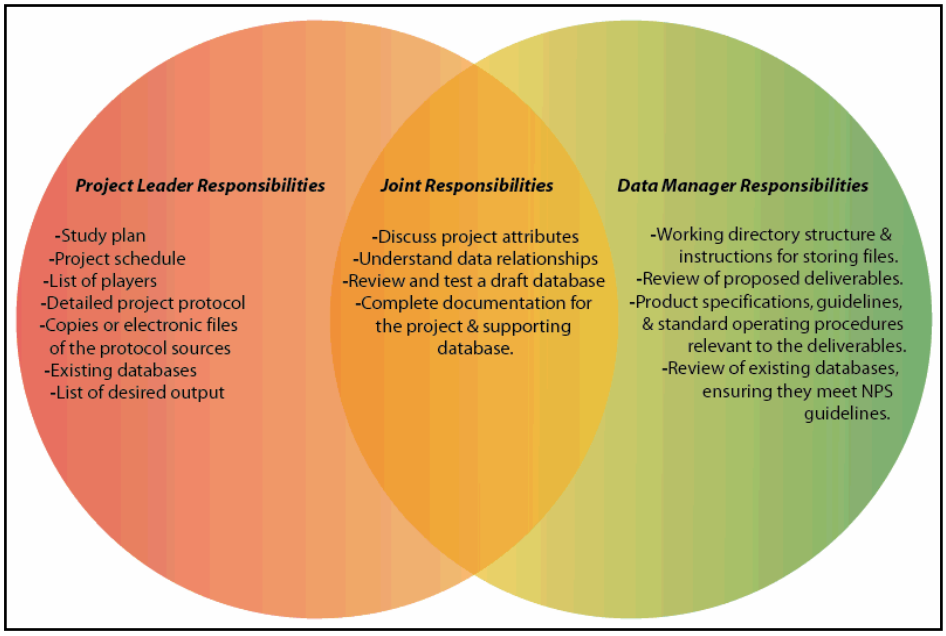


Figure 2.1. Joint Project and Data Management Responsibilities

Data managers

The data manager is responsible for ensuring the compatibility of project data with program standards, for designing the infrastructure for the project data, and for ensuring long-term data integrity, security, and availability. The data manager coordinates with the project leader throughout the project, as shown in Figure 2.1.

- Develop and maintain the infrastructure for metadata creation, project documentation, and project data management.
- Create and maintain project databases in accordance with best practices and current program standards.
- Provide training in the theory and practice of data management tailored to the needs of project personnel.
- Develop ways to improve the accessibility and transparency of digital data.
- Establish and implement procedures to protect sensitive data according to project needs.
- Collaborate with GIS specialists to integrate tabular data with geospatial data in a GIS system in a manner that meets project objectives.

Data managers will also work closely with the project leader to:

- Define the scope of the project data and create a data structure that meets project needs.
- Become familiar with how the data are collected, handled and used.

- Review quality control and quality assurance aspects of project protocols and standard procedure documentation.
- Identify elements that can be built into the database structure to facilitate quality control, such as required fields, range limits, pick lists and conditional validation rules.
- Create a user interface that streamlines the process of data entry, review, validation, and summarization that is consistent with the capabilities of the project staff.
- Develop automated database procedures to improve the efficiency of the data summarization and reporting process.
- Make sure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of the project data.
- Ensure regular archival of project materials.
- Inform project staff of changes and advances in data management practices.

GIS specialists

The GIS specialists manage spatial data themes associated with network inventory and monitoring projects, as well as other spatial data related to the full range of park resources. They incorporate spatial data into the GIS. They also maintain standards for geographic data and are responsible for sharing and disseminating GIS data throughout the network.

The GIS specialists will work in collaboration with project leaders to:

- Determine the GIS data and analysis needs for the project.
- Develop procedures for field collection of spatial data including the use of GPS and other spatial data collection techniques.
- Display, analyze, and create maps from spatial data to meet project objectives.
- Properly document data in compliance with spatial metadata standards.

GIS specialists will also work directly with data managers to:

- Design databases and other applications for the network.
- Create relationships between GIS and non-spatial data and create database and GIS applications to facilitate the integration and analysis of both spatial and non-spatial data.
- Establish and implement procedures to protect sensitive spatial data according to project needs.
- Develop and maintain an infrastructure for metadata creation and maintenance.
- Ensure that project metadata are created and comply with national and agency standards.

3. Data Management Process and Work Flow

The network is involved with short-term and long-term projects that share many work flow characteristics and generate data sets and other products that must be managed and made available to current and future users. Short-term projects usually last from one to three years and include individual research projects, inventories, or pilot work done in preparation for long-term monitoring. Long-term projects include the network's vital sign monitoring projects and other multi-year research and monitoring projects performed by other agencies and cooperators. Successful project management is based in part on establishing and communicating data stewardship responsibilities for all staff at all stages of a project.

3.1 Project Work Flow

The five stages of a project shown in Figure 3.1 involve the roles and responsibilities for project and data management discussed in this section.

Planning and approval

The project leader and program administrator are responsible for defining the scope and objective(s) of the project, establishing funding sources, and obtaining permits and compliance clearances. The data manager contributes where appropriate as data processing and reporting schedules are set and budgets for data management activities are discussed and incorporated.

Design and testing

The project leader is responsible for adopting or developing and testing project methodology to meet project objectives. The project leader and data manager work together to establish appropriate data management procedures, develop the data design, and prepare a data dictionary, which comprehensively lists and describes the database fields and data value domains. The data manager develops a new database application or adapts an existing database to meet project requirements.

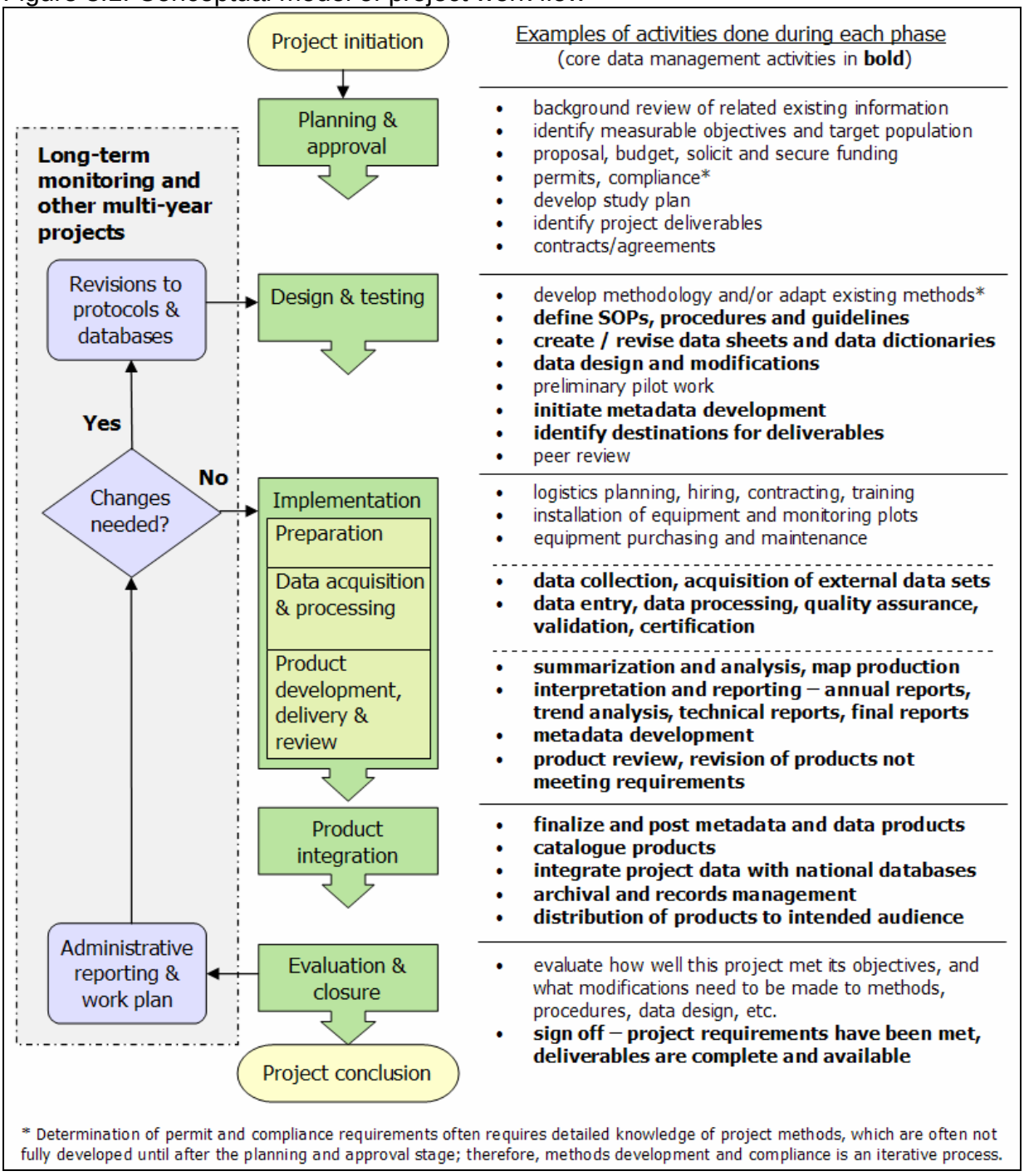
Implementation

During implementation data are acquired, processed, error-checked and documented. Reports, maps, GIS themes, and other products are developed and distributed. The project leader oversees all aspects of implementation, including logistics, contracting, training, and equipment procurement, data acquisition, report preparation and final delivery of products to resource managers. Data management staff provide training and support for database applications, GIS, GPS, and other data processing applications; facilitation of data summarization, validation and analysis; and assistance with the technical aspects of documentation and product development.

Product integration

The network maintains master copies and archives of all network-generated data on a local server. These are used to update NPS databases at appropriate intervals, e.g. annually updates to WRDSTORET for water quality data. Section 4.2.1 lists and describes the Service-wide databases used by the network for archival and distribution of data and information resources. The network also works with other partners to share and integrate results from network projects.

Figure 3.1. Conceptual model of project work flow



Evaluation and seasonal or final project wrap-up

For long-term monitoring and other cyclic projects, this phase occurs at the end of each field season as part of an annual review of the project. For non-cyclic projects, this phase represents the completion of the project. Upon season or project closure, project records are updated to reflect the status of the project and its associated deliverables in a network project tracking application. After products are catalogued and made available, program

administrators, project leaders, and data managers work together to assess how well the project met its objectives, and to determine what might be done to improve various aspects of the methodology, implementation, and formats of the resulting information. For monitoring protocols, careful documentation of all changes is required. Changes to methods and procedures are maintained in a tracking table associated with each document. Significant revisions, as defined in each protocol, may require additional peer review.

3.2 Data Life Cycle

Project data take on different forms and are maintained in different places as they are acquired, processed, documented, and archived. Figure 3.2 is a conceptual model that shows relationships between forms of data, databases, and data processes. Listed here are databases and forms of data:

- *Raw data* – Analog data recorded by hand on hard-copy forms and digital files from handheld computers, GPS receivers, telemetry data loggers, etc.
- *Working database* – A project-specific database for entering and processing data for the current season (or other logical period of time).
- *Certified data and metadata* – Completed data and documentation for short-term projects, or one season of completed data for long-term monitoring projects. Certification is a confirmation by the project leader that the data have passed all quality assurance requirements and are complete and ready for integration and distribution. Metadata records include the detailed information about project data needed for its proper use and interpretation (see Chapter 7).
- *Master database* – Project-specific database for storing the full project data set, used for viewing, summarizing, and analysis. Only used to store data that have passed all quality assurance steps.
- *Reports and data products* – Information that is derived from certified project data.
- *Edit log* – A means of tracking changes to certified data.
- *National databases and repositories* – Applications and repositories maintained at the national level, for the purpose of centralization, archival, and distribution among NPS units and with cooperators and the public.
- *Local archives and digital library* – Locally stored data, metadata and other products generated by projects.

The typical data process flow is described here and also shown in Figure 3.2:

1. *Acquire data* – For data recorded by hand in the field, data forms should be reviewed regularly (e.g., daily or weekly) for completeness and clarity in order to identify and resolve errors as close to their origin as possible.

2. *Archive raw data* – Copies of all raw data files are initially stored intact. Digital files are copied to the digital library section for the project; hard copy forms are either scanned and placed in the digital library, or are copied and placed in the archives. Archival or scanning of hard copy data forms may occur at the end of a season as a means of retaining all marks and edits made during the verification and validation steps. Once a project is completed or annual data results are validated and reported, original and intermediate data sets may be disposed in order to avoid unnecessary clutter in the information system. Disposing of data assets must be approved by at least the project

leader and the data manager based on established criteria for disposition that is documented in the data management procedures for the project.

3. *Data entry / import* – Analog data are entered manually, and digital data files are uploaded to the working database.

4. *Verification, processing and validation* – Verify accurate transcription of raw data and validate data using database queries to identify and resolve missing data values, out-of-range values, and logical errors.

5. *Documentation and certification* – Develop or update project metadata and certify the data set. Certification is a confirmation by the project leader that the data have passed all quality assurance requirements and are complete and documented. It also means that data and metadata are ready to be posted and delivered externally or used locally.

6. *Archive versioned data set* – Copies of the certified data and metadata are placed in the digital library. This can be accomplished by storing a compressed copy of the working database, or by exporting data to a more software-independent format (e.g., ASCII text; see Chapter 10).

7. *Post data and update national databases* – To make data available to others, certified data and metadata are posted to national repositories such as the NR-GIS Data Store and Biodiversity Data Store. In addition, national databases such as NPSpecies, NPSTORET, and NR-GIS Metadata Database are updated. Data and data products containing protected information are posted according to the guidelines and access levels outlines in chapter 9.

8. *Upload data* – Certified data are uploaded from the development database to the production database.

9. *Reporting and analysis* – Certified data are used to generate data products, analysis, and reports, including semi-automated annual summary reports for monitoring projects. Depending on project needs, data can be exported for analysis using other software applications and/or summarized within the database.

10. *Store products* – Reports and other data products are stored according to format and expected demand – either in the digital library, on off-line media, or in the document archives.

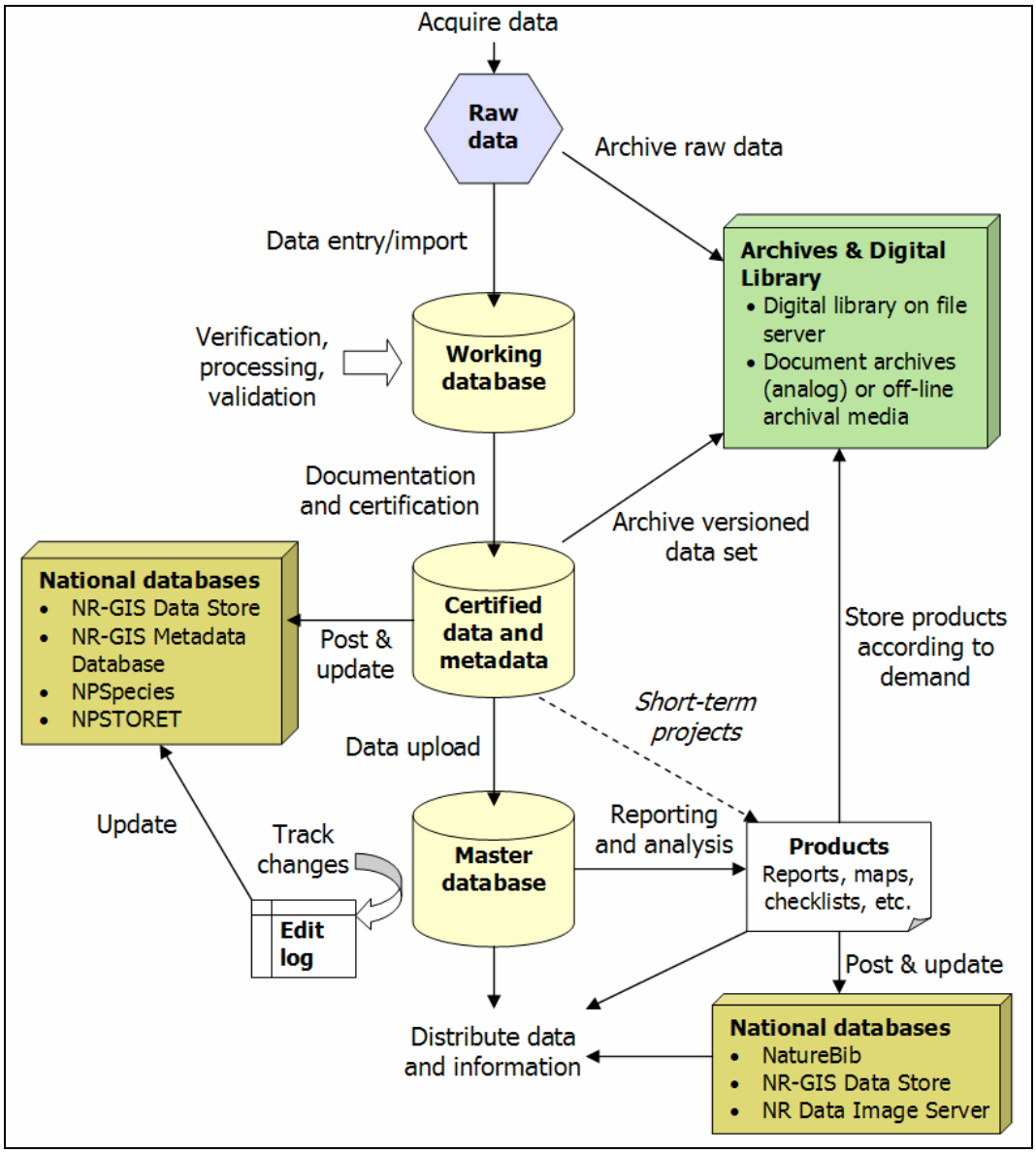
11. *Post products and update national databases* – To make products available to others, reports and other products are posted to national repositories such as NR-GIS Data Store or the NR Data Image Server. In addition, products are catalogued in NatureBib. Data products may not be posted if they contain protected information about the nature or location of rare, threatened or endangered species, or other natural resources of management concern (see Chapter 9).

12. *Distribute data and information* – Deliverables can be shared and distributed in a variety of ways - via the web-based national databases and repositories, by FTP or mailing in

response to specific requests, or by providing direct access to project records. Distribution is subject to legal requirements to protect information about sensitive resources (see Chapter 9).

13. *Track changes* – All subsequent changes to certified data are documented in an edit log, which accompanies project data and metadata upon distribution. Significant edits will trigger reposting of updated data and products to national databases and repositories.

Figure 3.2. Diagram of the typical project data life cycle.

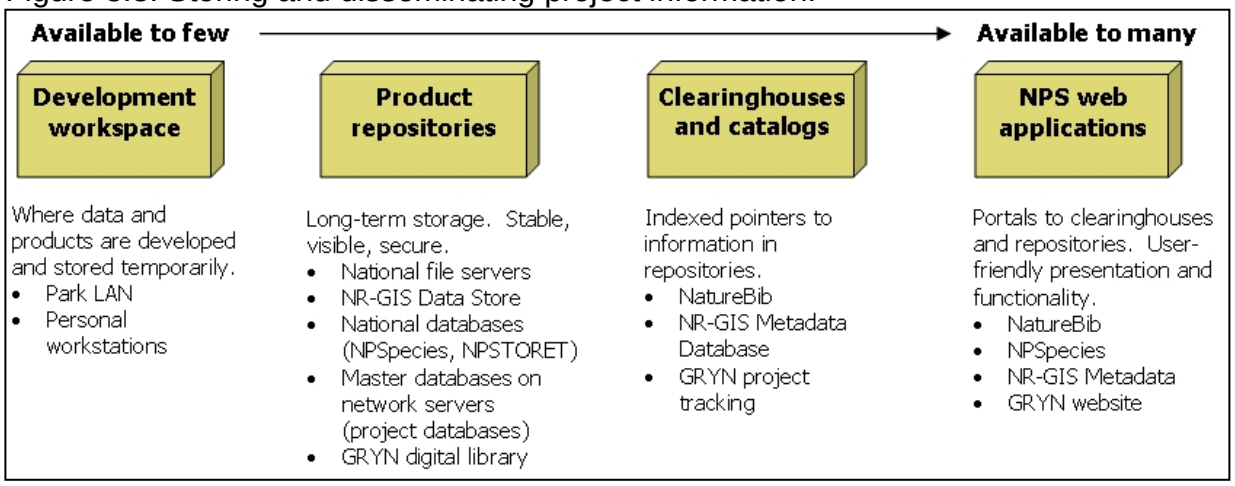


This data flow occurs iteratively (usually annually) for long-term monitoring projects and once for short-term projects. For projects spanning multiple years, decision points include whether or not a separate working database is desirable, and the extent to which product development and delivery is repeated year after year.

3.3 Integrating and Sharing Data Products

Final and seasonally certified project data and products need to be secured in long-term storage and made available. This requires using a variety of information systems such as product repositories, clearinghouses, and web applications. Each of these systems has a different purpose and function, as shown in Figure 3.3.

Figure 3.3. Storing and disseminating project information.



The specific repositories for most GRYN products are indicated in Table 3.1.

Table 3.1. Repositories for GRYN products.

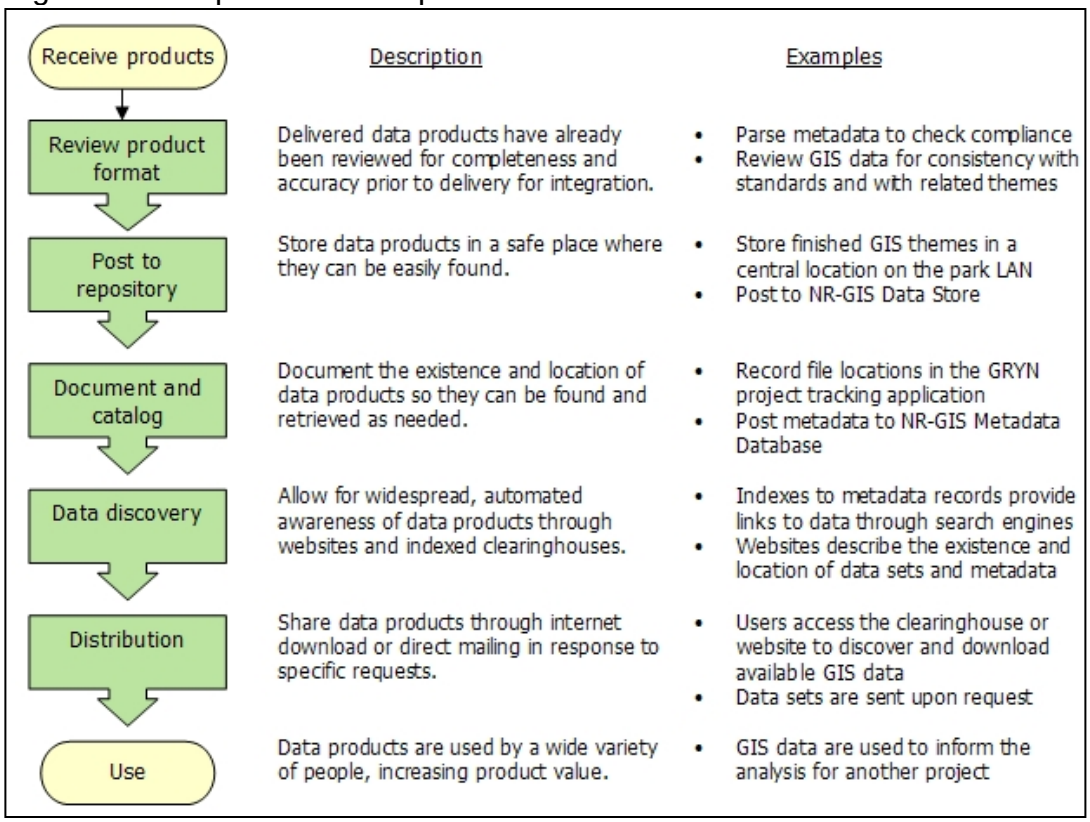
Item	Repository
Reports	GRYN digital library; posted to NR Data Image Server, linked and accessed through the catalogue record in NatureBib; Park collection (hard copy)
Digital data, metadata, and other products • Raw and processed data sets • Metadata, protocols, SOPs • Completed reports • Digital photographs	GRYN data servers, NR-GIS Data Store, Biodiversity Data Store, and other cooperators and agencies.
Project materials • Vouchers specimens, field data forms	GRYN and Park archives and collections, or another specified repository.
Administrative records	GRYN offices and/or park offices, park archives, National Archives

3.3.1 Data distribution

The process of product distribution involves several steps (Figure 3.4). In most cases the data manager or GIS specialist reviews products for conformance with format standards then transmits the products to the appropriate repository. All products are reviewed for completeness and accuracy by the appropriate resource specialist prior to delivery to the

data manager. Stored products are documented by posting metadata to the NPS clearinghouse and by updating records in the GRYN project tracking application. These metadata records provide contact and distribution information that help users obtain and use data sources and products.

Figure 3.4. Steps involved in product distribution.



Chapter Credits

This chapter was prepared by John Boetsch (NCCN) and adapted from concepts and material developed in collaboration with Dorothy Mortenson (SWAN), Velma Potash (CACO), Sara Stevens (NCBN), and Doug Wilder (CAKN). Special thanks to Dorothy Mortenson and Doug Wilder for significant contributions and ideas.

4. Data Management Infrastructure and Applications

The I&M Program relies on park, regional, and national IT personnel and resources to design and maintain a functional infrastructure to support I&M database and information system applications. This chapter briefly describes IT infrastructure and focuses on specific applications that are central to I&M network data management in a developing NPS Enterprise Architecture.

4.1 Network Computer Infrastructure

Due to the location of the GRYN offices separate from a park or other NPS office with conventional NPS computer network service, GRYN maintains a small client-server computer architecture that relies on disconnected data transfer to exchange information with the parks. The distributed hardware configuration shown in Figure 4.1 will suffice until a direct connection is established between park and network computers. The network manages common tables and high-value, long-term project databases within a system of revolving network attached storage devices to support a distributed, multi-user environment. There are four data servers that comprise the network infrastructure – one located at each of the three parks and one at the network's main office. These servers function as independent data nodes that can be accessed by park staff served by a local area network. They are also integrated in that common tables are replicated regularly among data nodes and network databases are replicated to each server.

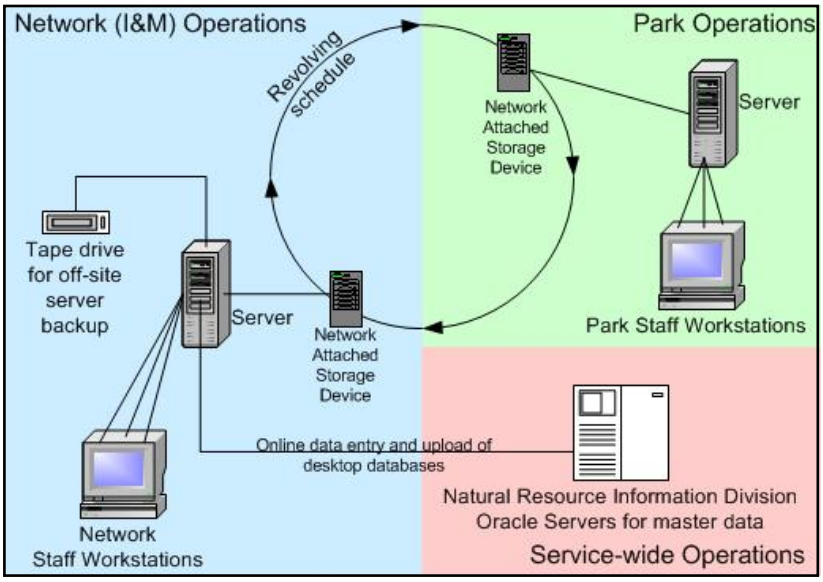


Figure 4.1. Network Computer Architecture

4.2 Network Information Management Systems

The network's information system strategy involves modular, standalone project databases that share design standards and use common data tables where possible, e.g. contacts and project personnel. Individual project databases are developed, maintained, and archived separately in Microsoft® Access and ESRI® GIS formats. Separate database applications allow greater flexibility in meeting the needs of each project but can be a barrier to synthesizing and consolidating data from related projects. Network and park data management and GIS staff work to develop resources, expertise, and institutional support for integrated and centralized databases that support park-level user requirements for natural resource management. Both the user requirements analyses and subsequent database development require a large initial investment followed by a long term commitment of resources to maintain, update, and support database applications. In order to be feasible,

investments made in complex interdisciplinary databases must be made with full participation from and understanding by park managers, scientists, and information technology staff. National level business requirements, such as summarizing and reporting core variables, must also be considered.

4.3 Information System infrastructure

The network stores data on national servers that host several master NPS applications, including NatureBib, NPSpecies, NR-GIS Metadata and Data Store. Centralized NPS repositories that include some of the network's data include the Biodiversity Data Store and the I&M Protocol Database. Public access sites to network data resources include current and planned portals to NatureBib, NPSpecies, NPSFocus, and the STORET National Data Warehouse.

Network and park data servers hold master network applications for data sets that are available locally for use by park and network natural resource specialists and managers. These include common tables, certified project data, a project tracking database, working data sets for NatureBib, NPSpecies, Dataset Catalog, and NPSTORET, digital libraries of developmental and production files and data sources, and downloads from national NPS or other data sources.

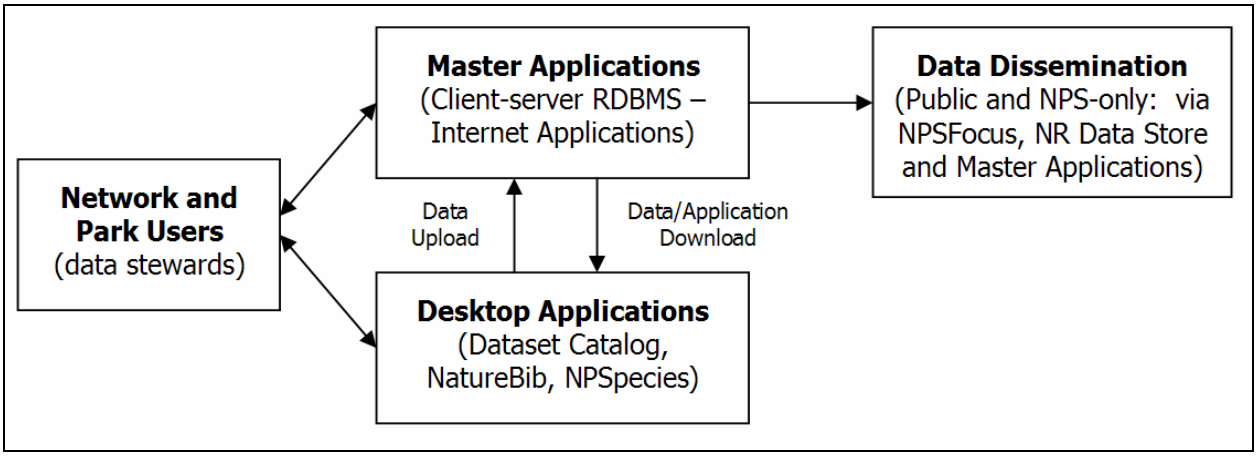
Each of the GRYN park servers and the network server uses a hierarchical folder and filing structure that meets the needs of local users. Files in development are commonly stored in separate folders from those used for production. The network's approach to folder and file management is explained in appendix B. Files for finished products are typically read-only. "Inbox" folders can be used as bins where users can save project-related files for processing and filing. Standards such as naming conventions and hierarchical filing are enforced within the production sections of the folder system. Although less stringent in other sections, these conventions are encouraged as good practice.

4.4 National Level I&M Database Applications

Integrating inventory and monitoring results at the network level involves national-level programmatic data and information management strategies that help provide consistent, long-term data and information sources for stewardship decisions, resource protection, and interpretation. The network maintains local access to all certified project data for use by park and network staff to help inform resource management and protection at parks. The network also provides data to the NPS Natural Resource Program Center (NRPC) and the I&M Program are involved in to centralized data archiving and distribution solutions developed by the NPS Natural Resource Program Center and the I&M Program. These databases and applications provide archive and accessibility for network data, and summary use by NPS and DOI managers and central office personnel to answer requests from Congress and for budget, program, and project planning.

To achieve an integrated information management system, three of the national-level data management applications (NatureBib, NPSpecies, and NR-GIS Metadata Database) utilize a distributed architecture (Figure 4.2) with master Oracle RDBMS applications available via web browser and desktop database applications for off-line data entry and access.

Figure 4.2. Model of the national-level application architecture.



NatureBib

NatureBib is the master database for bibliographic references that merges a number of previously separate NPS databases such as Whitetail Deer Management Bibliography (DeerBib), Geologic Resource Bibliography (GRBib), and others. It also contains citation data from independent databases like NPSpecies, Dataset Catalog, and NR-GIS Metadata Database. It currently focuses on natural resource references, but may eventually be linked to references on cultural resources and other park operations. As with NPSpecies and NR-GIS Metadata Database, it is possible to download data from the master web version into the Microsoft® Access desktop version that can be used locally on computers with limited internet connectivity *(http://www.nature.nps.gov/nrbib)*.

NPSpecies

NPSpecies is the master species database for the NPS. The database lists the species that occur in or near each park, and the physical or written evidence for the occurrence of the species (e.g., references, vouchers, and observations). Taxonomy and nomenclature are based on the interagency Integrated Taxonomic Information System (ITIS). The master version of NPSpecies for each park or network can be downloaded from the master website into an Microsoft® Access version of NPSpecies. The internet-based version is the master database, which can be accessed via password-protected logins administered by park, network and regional stewards assigned for each park and network. The master database requires certification of species lists by networks before any data will be available to the public. NPSpecies is linked to NatureBib for bibliographic references that provide written evidence of species occurrence in a park and will be linked to NR-GIS Metadata Database to document biological inventory products. The Microsoft® Access application and additional details can be found at the NPSpecies website *(http://science.nature.nps.gov/im/apps/npspp/index.htm)*.

Dataset Catalog and NR-GIS Metadata Database

Dataset Catalog is a desktop metadata database application developed by the I&M Program to provide a tool that parks, networks, and cooperators can use to inventory and manage data set holdings. Although not designed as a comprehensive metadata tool, the Dataset Catalog is used for cataloging abbreviated metadata about a variety of digital and non-digital natural resource data sets. The Dataset Catalog helps parks and networks begin to meet

Executive Order 12906 mandating federal agencies to document all data collected after January 1995. It provides brief metadata and a comprehensive list about all resource data sets for use in data management, project planning, and more stringent metadata activities. As with other service-wide applications, the master metadata database (NR-GIS Metadata Database) is available through a website and will be linked to NPSpecies (the NPS species database) and NatureBib (the bibliographic database). It will be possible to download a version in Microsoft® Access format from the master website (*Dataset Catalog: http://science.nature.nps.gov/im/apps/datacat/index.htm and NR-GIS Metadata Database: http://science.nature.nps.gov/nrdata).*

NPSTORET

NPSTORET is the NPS desktop database designed to facilitate standardized park-level reporting for STORET, an interagency water quality database developed and supported by the Environmental Protection Agency's (EPA). STORET is an electronic warehouse for local, state, and federal water quality data collected in support of managing the nation's water resources under the Clean Water Act. STORET is used by NPS as a repository of physical, chemical, biological, and other monitoring data collected in and around national park units by park staff, contractors, and cooperators. The NPS Water Resource Division (WRD) operates WRD STORET and makes periodic uploads to the EPA STORET National Data Warehouse so that data collected by and for parks will be publicly accessible. NPS Director's Order 77 instructs the NPS to archive water quality data in STORET. Network staff and cooperators will use NPSTORET for data entry and management and to prepare data for upload to STORET by WRD staff (Figure 4.3). To support spatial analysis and functionality NPSTORET includes a field for the National Hydrography Dataset (NHD) reach code where water samples are collected. This allows sampling stations to be indexed to hydrographic features and analyzed with other resource layers in a GIS using a suite of hydrographic network and hydrologic tools. The Microsoft® Access application and additional details can be found at: *http://www.nature.nps.gov/water/infodata.htm.* Additional information on STORET can be found at: *http://www.epa.gov/storet.*

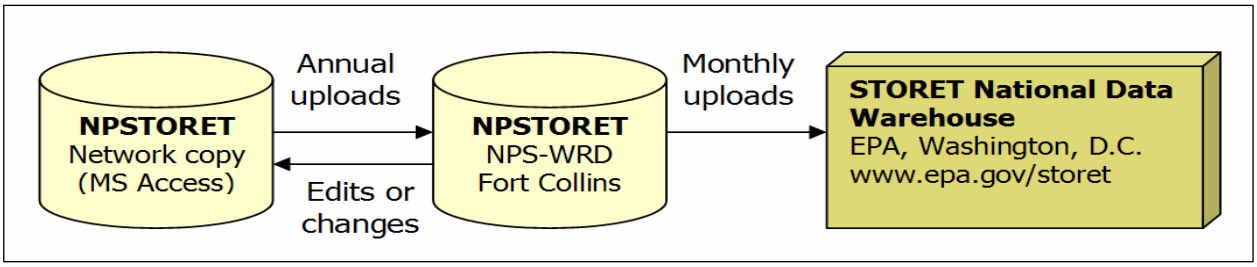


Figure 4.3. Water Quality Data Flow

Natural Resource Database Template

The Natural Resource Database Template (NRDT) is a flexible, relational database in Microsoft® Access for storing inventory and monitoring data (including raw data collected during field studies). This relational database can be used as a standalone database or in conjunction with the GIS software (e.g., ArcView™ or ArcGIS™) to enter, store, retrieve, and otherwise manage natural resource information. The template has a core database structure that can be modified and extended by different parks and networks depending on the components of their inventory and monitoring projects and the specific sampling protocols in use. NRDT is designed to support some standardization among monitoring protocols.

Approved monitoring protocols, including the databases that use the NRDT are available through a web-based protocol clearinghouse (see next item). A description of the Database Template application, a data dictionary, and example implementations are located on the NR Database Template website *(http://science.nature.nps.gov/im/apps/template/index.htm).*

Natural Resource Monitoring Protocols Clearinghouse

The Natural Resource Monitoring Protocol Clearinghouse (i.e., Protocol Database) is a web-based clearinghouse of sampling protocols used in national parks to monitor the condition of selected natural resources. The clearinghouse provides a summary of each monitoring protocol and links to download full protocols and database designs. See the Protocol Database website for available protocols. (*http://science.nature.nps.gov/im/monitor/protocoldb.cfm*).

NR-GIS Data Store

The NR-GIS Data Store is a key component of the data distribution strategy used by the I&M Program. The NR-GIS Data Store is a graphical search interface that links data set metadata to a searchable data server on which data sets are organized by NPS units, offices and programs. The interface allows customized public or protected searches of natural resource data sets, inventory products, and GIS data produced by the I&M and Natural Resource GIS programs. Each park or network is able to post and manage its data on the server. The NR-GIS Data Store will be integrated with the master NR-GIS Metadata Database application to streamline programmatic data documentation and distribution processes. The simple browse function of this server can be accessed at: *http://nrdata.nps.gov/*.
See the NR-GIS Data Store website for further information (*http://science.nature.nps.gov/nrdata*).

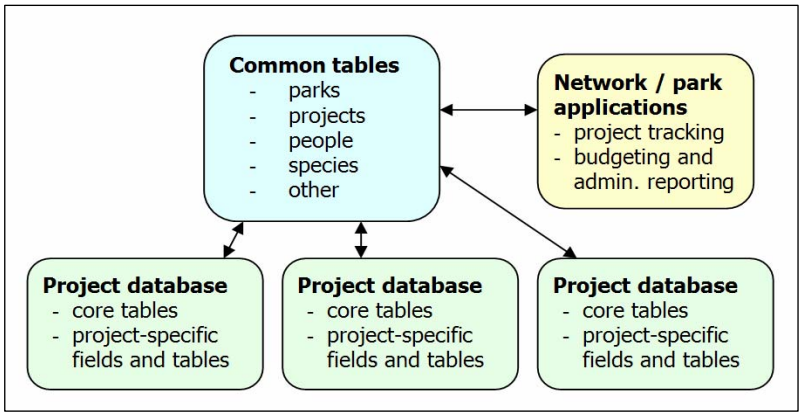
NPS Biodiversity Data Store

The Biodiversity Data Store (BDS) is a digital repository of documents, GIS maps, and data sets that contribute to the knowledge of biodiversity in National Park units, including species occurrence, distribution and abundance. The web site includes public and secure areas for searching and downloading data. The network will work to understand and manage the relationship between the BDS and the NPS GIS Data Store.

4.5 Project Database Standards

To promote compatibility among data sets that will be used in summary analysis and reporting across vital signs, network databases contain ***common lookup tables*** (Figure 4.4), ***core project tables*** standardized with regard to data types, field names, and domain ranges, and ***project-specific tables***. The network adapts or develops databases

Figure 4.4. Common lookup tables and satellite databases



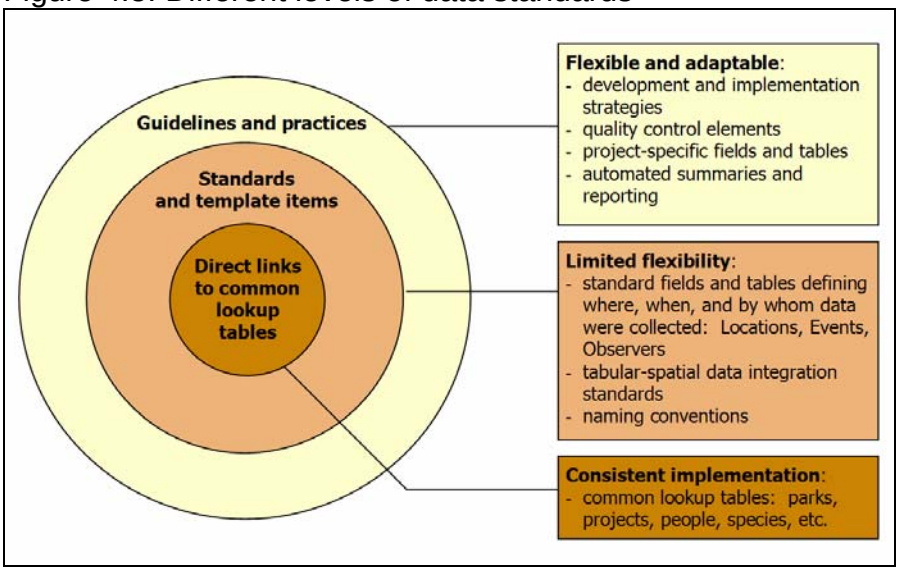
to be compatible with those maintained by other networks and cooperators managing similar data sets.

In the future a client-server database system may be developed at the network to associate common database components that are replicated among data nodes hosted at each network park. Individual project databases will continue to access these common tables.

4.5.1 Levels of data standards

Because common lookup tables are stored in one place and are referred to by multiple databases, they represent the highest level of data standard because they are implemented identically among data sets. The second level of standards is implied by core template fields and tables, which are standardized where possible, but project-specific objectives and needs could lead to varied implementations among projects. The third level of standards is applied most flexibly to accommodate the range of needs and possibilities for each project, yet always with compatibility and integrity in mind. Figure 4.5 presents the resulting variation in implementation of these differing levels as a "bull's eye", with the common lookup tables providing the most consistent implementation and hence the smallest range of variation.

Figure 4.5. Different levels of data standards



4.5.2 Project tracking application

To support program coordination and annual reporting, and to improve accountability for the products of natural resource inventory and monitoring efforts, the network will use a project tracking database. The primary functions of this application include:

1. Maintaining the list of projects – By maintaining a single list of natural resource data projects, it is much easier to quickly find project-related information (e.g., status, funding sources and amounts, objectives, contact information) and summarize that information for administrative reports.
2. Tracking product deliverables – For each project a comprehensive list is maintained of what deliverables are expected and when. Once they are delivered and posted or archived, this function shifts to being a finding aid for available products. Deliverables

are first specified at project initiation and information is updated at various project milestones (e.g., contracting, product delivery, archival).

3. Managing project codes – These are intelligent alphanumeric codes used to tie together digital information in various, minimally connected systems (e.g., RPRS, PMIS), along with analog materials that cannot otherwise be linked to an integrated information system. These codes are also used to link to data in databases and GIS themes, especially where information from multiple sources is stored together.

The master version of this application will be maintained at the network office and made available to each park. Project leaders, the network data manager, GIS staff, and the network coordinator share responsibility for the maintenance and availability of the project tracking data and reports. Each of these staff will be able to make certain changes to update information about project status, deliverable details, etc. Database views will be created to help project leaders keep on schedule, and to facilitate quick reporting on project status, accomplishments and delivered products.

Chapter Credits

This chapter was adapted from material prepared by John Boetsch (NCCN) and Patrick Flaherty (APHN). Section 4.2 - National Information Management Systems was adapted from material written by Lisa Nelson (WASO).

5. Data Acquisition and Processing

The network's peer reviewed and NPS-approved monitoring protocols list specific ecosystem parameters that are collected by observation, measurement, and/or prediction and stored as data. To be useful to park managers these data must be processed, analyzed, mapped, displayed, and reported not only as stand alone vital signs but in ways that demonstrate the dynamic connections among and between vital signs. Some of the vital signs selected by network parks are monitored using methods developed before the GRYN was established. Data from these past and continuing efforts contribute to scientific understanding and park stewardship regardless of their basis in local methods or peer-reviewed I&M protocols. The network engages natural resource professionals at all organizational levels to help determine the relevance to monitoring objectives for a variety of park and other data sources. Data that are certified by subject matter experts as valid, complete, and fully documented with FGDC and NPS spatial and biological metadata are processed to contribute to the network's inventory and monitoring objectives.

Table 5.1. Abbreviated Data Development Model

1. Identify issues and concerns
2. Define the purpose and need for data collection and analysis
3. Develop explicit monitoring objectives or inventory criteria (these are key questions addressing the issue or concern within the scope of the purpose and need)
4. List measurable, observable, and predictable variables associated with each key question
5. Formalize (document) information needs
6. Develop a Data Dictionary for field names, lists of values, quality factors, and metadata characteristics
7. Select or develop an appropriate sample design
8. Identify and assign explicit data stewardship roles and responsibilities
9. Write a complete monitoring protocol or inventory study plan
10. Design or adopt/adapt a database (including quality control elements)

1. Plan for data acquisition (**beginning of data life cycle)**
2. Collect data – field and office components
3. Process data (includes verification, transfer, addition of required attributes)
4. Store, organize, and secure data
5. Use, analyze, and report data
6. Maintain and serve data and derived products
7. Archive data (long term storage that may require media and/or platform transfer)
8. Dispose of data that are no longer needed, if appropriate (**end of data life cycle**)

Newly acquired vital sign data funded by the network involve the steps outlined in Table 5.1 and will meet the following requirements:

- Base data acquisition on documented and approved monitoring protocols
- Adhere to a data dictionary approved by the project leader and the network data manager
- Use common and/or existing data formats and field definitions to promote crosswalk and comparison with data from other times and places.
- Enter data in a database approved by the network data manager
- Provide data verification in the field and in the office by field crew members and the field crew leader
- Process the data with quality control features built in to a database application

5.1 Database Design

To promote data comparability among parks and networks GRYN draws on and extends the concepts and structure of the NPS Natural Resource Database Template and the NPS conventions for naming database objects. It is also important for GRYN to consider design elements that promote data comparability with other land and resource management entities in the Greater Yellowstone Ecosystem. Before starting a new database design from scratch the network will look to the following sources for existing designs and database objects:

- NPS Natural Resource Database Template
- I&M Protocol clearinghouse
- Other NPS I&M networks
- Cooperating agencies and entities
- Other parks
- Other agencies

In particular the network will review and associate, where appropriate, existing database and application design elements from the US Forest Service because their data structures store ecosystem parameters for administered lands surrounding Yellowstone National Park and Grand Teton National Park. Achieving some consistency in common data objects between the parks and the surrounding forests promotes ecosystem-based analysis and reporting.

5.2 Data Acquisition

New data collected by the network are from natural resources inventories or vital sign monitoring projects and collaborative efforts between the network and other regional programs or institutions. Data for these projects will be collected by I&M personnel, park staff, partners, and cooperators. Natural resource inventories help to identify the presence or occurrence, and status of park natural resources and represent an important component of park stewardship. A listing, description, and park status of standard NPS Natural Resource Inventories conducted by the Inventory and Monitoring Program are available online at http://science.nature.nps.gov/im/inventories.htm. Network-base inventory results are available via the GRYN web site (see title page for web address). Monitoring data are collected and processed according to requirements specified in peer reviewed and NPS-approved GRYN Vital Sign monitoring protocols. Vital signs are a subset of physical, chemical, and biological elements and processes of park ecosystems that are selected to represent the overall health or condition of a park's natural resources, known or hypothesized effects of stressors, or elements that have important human values. Information about the status and trend of vital signs contributes to the decision making processes used by park managers. Chapter III of the network's Vital Signs Monitoring Plan lists the vital signs and explains the selection process.

5.2.1 Data Discovery and Data Mining

Data discovery in the Greater Yellowstone Network refers to the process of searching out existing natural resource data and information that are useful to the I&M Program mission. Since 2000 GRYN has identified, catalogued, and entered thousands of data records and data sources in Dataset Catalog, NatureBib, and NPSpecies. The bulk of this effort occurred during the inventory phase of network development from 2000 through 2004 (GRYN 2000). These data discovery efforts primarily involved interviews with park natural resource specialists and targeted searches through hardcopy and electronic data sources stored at

park service units, other government offices, academic institutions, and collection facilities such as libraries, museums, and herbaria. Those efforts represent a different approach than more fundamental IT industry 'data mining' procedures that incorporate data input, data analysis, and results output using complex computer platforms and software applications for extracting knowledge from large volumes of data (US GAO 2004). The network does not intend to perform that type of data mining in the foreseeable future, but will continue to intentionally and incidentally identify and process relevant natural resource data sources related to GRYN parks and inventory and monitoring activities. A useful discussion and set of procedures for data mining activities that emphasizes a data catalog, bibliography, and metadata is available from the Klamath Network of the NPS I&M Program (Smith and Truitt, 2004). Additional background and guidelines for data mining in the NPS I&M Program are available on the npsweb: (http://www1.nrintra.nps.gov/im/datamgmt/DataMining/datamining.htm).

GRYN staff will find and review different sources for various types of data at the onset of new projects or during the development of new protocols as part of research for background information. Network staff use the following data sources via the internet and visits to local research or academic institutions, museums, and local parks.

Bibliographic/Literature
- National NPS Databases (e.g., NatureBib)
- Online literature databases (e.g., First Search or Biosis)
- Local document library
- Library catalogs (e.g., academic or research institutions)
- Park and other archives

Geographic Data
- Regional GIS Specialists
- Park GIS Specialists
- Federal and State Geographic data clearinghouses
- Local, state, and federal government offices

Biological/Natural Resources Data
- Automated National Catalog System (ANCS+)
- NPSpecies
- Voucher collections (museums, parks, universities)
- Network Parks
- Local, state, and federal government offices

Relevant information collected during a data discovery process is maintained at the network and/or at one or more network parks either electronically or in hard copy format depending on how it was collected. Geographic data sets collected during this process are documented with FGDC compliant metadata when possible and/or documented with brief metadata in Dataset Catalog.

4. **Dataset Catalog** – new data sets (spatial or non-spatial) are entered into Dataset Catalog.

- **Scientific citations** – The network uses the online and desktop versions of NatureBib.
- **NPSpecies** – Information relating to the biodiversity of network parks is entered into NPSpecies and linked to the available reference documents in NatureBib.
- **Reference Cabinets** – Hard copy materials are stored that relate to the network parks and the GRYN projects. All original data sheets, final reports, and contracts are stored in a fire proof cabinet and/or in more than one building as insurance against total loss. General references as well as those references linked to NPSpecies records in NatureBib are stored in file cabinets in the GRYN office.

Information collected during the data discovery process is likely to be ***legacy data***, which may be of high value to the network parks but differs in structure and intent from current NPS, I&M Program, or network data models, and/or is based on a different protocol with different objectives and methodology. Legacy data resources are primarily from inventory and monitoring efforts that originated before the NPS I&M Program. GRYN regards all legacy data sources as potentially useful. Network staff will coordinate with park staff to minimally document (e.g. in Dataset Catalog) those legacy data sources discovered incidentally during normal park and network business. Due to the volume of legacy data sources, particularly at YELL, investments in additional or extensive data discovery efforts must be reviewed and approved by project leaders and resource specialists to determine if candidate data sources can be documented as valid and related to the objectives of the vital signs monitoring program. If so, then legacy data may be converted using I&M funds to file formats compatible with the current standards. The network may scan hardcopy references and materials, saving them as PDF files in a digital library and transferring a copy to the NatureBib document center or NPS Biodiversity Data Store.

5.2.2 Field Studies

Biological inventories and monitoring projects are the most common examples of field studies conducted by the network. All NPS employees, volunteers, educational partners, contractors, etc. must conform to the National Park Service Geographic Information System Data Specification for Resource Mapping, Inventories, and Studies. (http://imgis.nps.gov/contract_specs.html) Following this guidance helps ensure that NPS investments result in usable GIS data in standard formats. The data manager is responsible for ensuring that data collection, data entry, verification, documentation, storage, and archiving for all field projects are consistent with the program and network standards. In addition to general operating procedures that define network-wide requirements, protocol specific standard operation procedures (SOP's) are developed for each protocol. The data manager works closely with the project leaders and network staff to develop these guidelines and methodologies for data collection, processing, and management. These activities range from specifying the proper design and use of hardcopy and electronic data entry forms to outlining calibration procedures for automated data loggers. Refer to the individual vital sign monitoring protocols for specific data management procedures. General NPS examples of protocol specific SOP's are available online at http://www1.nrintra.nps.gov/im/datamgmt/dmplanning.htm.

Field Data Acquisition Tools:

Since different protocols require unique data collection methods, the details about specific field data collection tools for individual projects are documented in each protocol's operating procedures. The network invests in technology based on the particular needs of each project rather than the allure of the latest, greatest gadgetry. The process to select equipment and methods will consider short term and long term cost, functionality, ease of use, user comfort levels, and support and training requirements. The network expects to use many of the general tools listed below.

Field Forms – this common method of recording field data is straightforward and inexpensive. Special attention must be paid to maximize clarity of written data values and minimize data form coding errors during the data recording process. Field forms require neat, legible handwriting and rigorous quality control. For some projects the network may transition from using hardcopy forms as the primary recording medium to electronic equipment, and continue to use field forms for backup when electronic equipment fails or is unavailable.

Field Computers – Collecting data in digital format in the field can increase data entry and processing efficiency and help reduce errors. Data can be directly transferred from the field computers to the office desktop computers to eliminate a data entry step. Data entry errors can be minimized by applying quality control functions in the electronic data entry forms. Field computers can complement hardcopy field notes, and hardcopy data sheets should always be available in case the electronic equipment is damaged, lost, out of power, or otherwise unavailable.

- **Handheld computers** (Personal Digital Assistants (PDA's), Portable Data Recorders (PDR's), etc.) **–** The small size and relative low cost of these devices make them worth considering for collecting field data. Those that are made for dusty, wet, and extreme temperature conditions or those that can be protected within a rugged container may be a good option for small field projects. However some may not have the battery, memory, or functional capacity for large data intensive field projects. Many rugged handheld computers are available with global positioning system (GPS) functions, as well as feature attribute data collection capacity. Most use Microsoft PocketPC or Palm operating systems which may require additional expertise and training in processing and programming to create the database structure for the handheld computer and transfer database content between computers.

- **Tablet PC's –** These have many of the same properties as laptop computers and provide the user with the convenience of a touch screen interface. They are bulkier, more expensive and harder to make rugged for field use than the handheld computers but are more powerful. Tablet PC's may be a better choice for field projects that are very data intensive. Because these units run Windows XP (Tablet Edition) the project database can usually be directly transferred from desktop units to field units without additional programming steps. The visibility of the computer screen in bright sunlight conditions is an important consideration for tablet PC's and handheld computers.

Automated Data Loggers – These are mainly used to collect physical environment data such as weather data or water quality parameter values. Data loggers require proper and

regular calibration and maintenance, so field crews must receive appropriate training, and standard operating procedures must include specific calibration and maintenance procedures. The resources required to set up, maintain, and manage automated data loggers should be carefully assessed and compared to alternative sampling techniques and equipment.

- **Permanently deployed devices –** Data from these devices must be retrieved and batteries changed on a regular basis. These intervals should be defined in the protocol and costs of visits and maintenance should be considered in the project budget.

- **Portable hand-held devices –** Deployed for sampling only during site visits, these are generally less expensive than units that are permanently deployed in the field.

Global Positioning System (GPS) Units – Two types of GPS units are typically used during field work to display and/or collect location information. The decision about which type to use depends on the requirements of the project related to the capabilities and limitations of the GPS equipment. See Appendix D for the GPS standard operating procedure.

- **Recreational GPS Units –** These are preferred mainly for all-purpose navigation rather than data collection because they store little or no attribute data and only general position information. Since most units do not record the accuracy of each position or feature, and the positions cannot be differentially corrected to improve the accuracy, recreational GPS units are not recommended for protocols or projects that require highly accurate location information. In some cases it is appropriate to use recreational GPS units to collect point features using position averaging to improve accuracy. Collecting line and area features with recreational GPS units is not recommended. Users are advised to record error estimates along with the coordinates.

- **Mapping GPS Receivers –** When used by well trained crew members, these receivers can collect highly accurate (sub-meter) location information, when using real time or post processed differential correction. Mapping grade receivers also support data entry forms for collecting feature attribute data that facilitates transfer to a GIS. Use of mapping grade units is encouraged for all line and area feature GPS data collection.

5.3 Data Entry

After data is recorded using the tools listed above, the following data entry procedures are necessary to maintain data flow and ensure data quality.

- Field crew members or other project staff will enter all data into a database approved by the project leader and network data manager.
- Project personnel will periodically transfer data files to the data manager. Individual protocols state the requirements for data transfer, backup, and archiving.
- The data manager maintains and updates the network's master copy of the database with data received from the field crews.

- All data is subject to quality assurance procedures. Chapter 6 discusses data verification and validation. Specific quality control activities are specified in each vital sign monitoring protocol.

Network data is stored and processed in local and national databases. Microsoft® Access databases and personal geodatabases are developed using the NPS Natural Resources Database Template. The flexibility of the NRDT allows customized local database applications to meet the needs and requirements of each project while using common tables for data values that repeat and are shared between projects (places, people, species, etc). Network database applications incorporate quality assurance mechanisms such as pick lists and validation rules. The network also processes its data using Service-wide data stores (listed in chapter 4) for centralized archival and distribution. The network keeps master data synchronized between local and Service-wide data stores.

5.4 Changes to Data Collection Procedures/Protocols

Changes to data collection procedures are made based on valid reasons for altering the methodologies. Most issues will be identified during the design and testing stages of the project and changes will be implemented prior to the collection of field data. The protocol development process includes attempting to identify and address any foreseeable issues that might occur with data acquisition and processing. Unforeseen issues may arise after data collection has begun which require changes to procedures or protocols. Significant changes to the protocols must be approved by the project leader, an I&M Program official at the Washington Office level, and the data manager. The I&M official must evaluate the proposed changes and determine if additional peer review is required before accepting the modifications.

Changes to protocols and associated data collection procedures may also occur as a result of scheduled program reviews. During the review, data are analyzed to determine if the current protocol is meeting stated objectives. If it is determined that the protocol has not achieved the desired results then recommended changes should be made.

5.5 Data Sources

A large percentage of data collected in network parks are collected by park personnel involved in projects initiated at the individual park level or by other NPS regional or national programs. The resulting data and associated products provide a great deal of information about park natural resources and are often relevant to the mission of the I&M Program.

5.5.1 Park Data

Network parks often use base funding or receive funding through other sources and programs to support park-level projects.

- **Park-based biological inventories** - network parks conducting their own park-based inventory projects generate data that can be used to supplement network-level inventories conducted by the I&M Program.
- **Park-based monitoring projects** - Park-level monitoring projects (such as vegetation and water quality) produce information that is valuable when developing network-level monitoring protocols.

- **Park and multi-park based projects** - other studies or projects conducted at the park or regional level that do not fall into one of the previous two categories (e.g., restoration projects).

5.5.2 Regional and National Programs

NPS regional and national programs support all of the parks within the network and also provide a good resource for natural resource information.

- **Air –** The NPS Air Resources Division (ARD) presents interpolated spatial and tabular air quality data for parks via the online Air Atlas: (http://www2.nature.nps.gov/air/). The NPS ARD compiles and analyzes air-related data from several national air monitoring networks, and performs broad-scale trend analyses relevant to GRYN air quality issues. The GRYN I&M Program will rely in part on these national scale monitoring networks to obtain trends for air-related vital signs.

- **EPMT -** Exotic Plant Management Teams (EPMT) collect and maintain data regarding the presence of exotic species in regional parks and the methods used to treat these species. This information is stored in the Alien Plant Control and Monitoring Database (APCAM) which is maintained by the EPMT data manager or EPMT liaison. http://www1.nrintra.nps.gov/BRMD/invasivespecies/exoticplants/index.htm

- **Fire Program –** data concerning the occurrence of fires within the network are maintained at the parks and regionally. National databases such as Fire-Pro, SACS and the soon-to-be implemented Fire Program Analysis (FPA) package (http://fpa.nifc.gov/) provide information regarding resources devoted to fires as well as fire occurrences. The NPS is also involved in efforts such as the Joint Fire Science Program (http://jfsp.nifc.gov/) that provides scientific information and support for fuel and fire management programs.

- **GIS –** The NPS OCIO GIS Division, the Natural Resource Program Center, and regional GIS staff provide guidance and assistance to network and park staff.

- **Water -** nearly all of the field data collected by the regional water resources program supports the water resources vital signs monitoring projects. The program also synthesizes, analyzes, and interprets water resources data collected by parks. The National Hydrography Dataset (NHD) is the NPS standard for spatial hydrography data. NPS Water Resource Division assists with water quality data management by providing NPSTORET and data transfer services for STORET.

- **Wildlife Management -** The NPS Wildlife Management Program is involved with coordinating long-term monitoring and assessment of wildlife populations.

5.5.3 Processing National Park Service Data

The I&M Program's current framework for natural resource information management (Figure 5.1) aims at achieving maximum return on investments made in data gathering, such that

relevant data and resulting information is available long term to multiple levels of the NPS organization (park, network, regional, national). The framework includes these elements:

- Provide spatial and attribute data standards and production methods for natural resource inventories
- Develop and support service-wide online natural resource database applications in Oracle RDBMS
- Provide desktop database applications that mirror master databases and promote standard data entry and organization
- Recommend a natural resource database template that allows local flexibility but also promotes design consistency for the purpose of sharing database designs and content
- Direct networks to hire data management staff and emphasize data management
- Mandate written network data management plans
- Require written data management procedures and responsibilities in inventory study plans and vital sign monitoring protocols

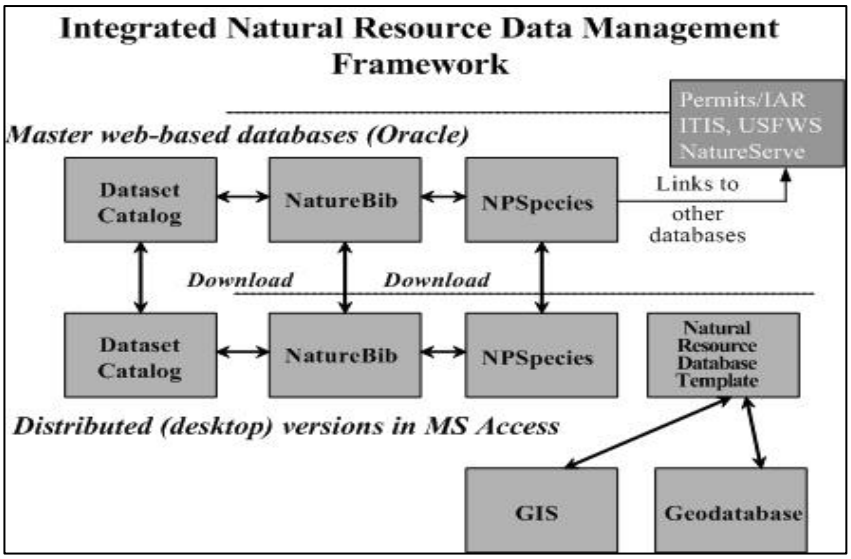


Figure 5.1. Natural Resource Data Management Framework

It is important that park, regional and network staff work closely together to ensure that information is maintained in a manner that promotes data sharing. Accordingly, the network data manager will:

- Coordinate with park and regional personnel to ensure that high quality data are available.
- Arrange for and/or provide training to park staff interested in learning to use NPSpecies, NatureBib, NPSTORET, and the Natural Resource Database Template.
- Ensure that information collected/maintained by the parks is entered into NPSpecies or NatureBib.
- Provide assistance with processing and storing voucher specimens collected in network parks.
- Assist with developing user requirements, database designs, and database application development based on the NPS Natural Resource Database Template and geodatabase.

Basic data processing steps include those listed here:

- Enter all new park biological inventory and monitroing data into NPSpecies (this is important for network-based and park-based biological inventories) and enter all associated references into NatureBib.
- Enter all park-based or regional reports and publications into the park library. Hard copies are stored in the appropriate file cabinets and electronic copies are archived in the proper directory on the park and network file server.
- Ensure that all GIS data is in the proper projection and accompanied by metadata that is complete and compliant with NPS format requirements.
- All data sets are entered into and tracked using Dataset Catalog.

5.5.4 External Data

Non-NPS and non-I&M data sources contribute to the development of the network's long term monitoring program. Certain vital signs may be addressed primarily by data that is collected and managed by non-NPS agencies or entities. For example, some vital sign parameters related to air quality, geology, and water can be referenced from existing non-NPS or non-I&M repositories and data stores. The agencies or organizations that collect these data have the expertise to conduct appropriate quality control procedures, and the capability to function as a repository and clearinghouse for the validated data. When necessary for analysis or summary reporting, the network acquires and processes off-site data according to the monitoring protocols. These "data snapshots" are archived by the network in cases where it may not be possible to repeat the same data acquisition and processing steps.

5.5.5 Processing External Data

Much of the data collected from external sources must undergo the following processing to meet the standards of the GRYN I&M Program.

- All GIS data obtained from other entities are stored in a format compatible with NPS information systems, has the correct spatial reference information, and has FGDC compliant metadata.
- All park-related biodiversity data obtained or received from other entities is entered into NPSpecies.
- Citations are entered into NatureBib for data taken from a report or published document.
- All reference materials obtained are kept in the network's library of electronic and/or hardcopy information resources.
- All data sets are entered into and tracked using Dataset Catalog.

Certain data sets will require more than the basic processing steps described above. The level of data processing required for external data sets such as those used in the vital signs monitoring program depends on the desired output. For example, if basic trends are of interest for air quality then relatively little processing is required. Where it meets the needs of the parks, the network will rely on trend analyses by other national programs. However, if more intensive analyses are needed, such as conducting in-depth analyses at a specific location rather than monitoring regional trends, the network will perform additional processing. In such cases, the specific protocols provide the necessary data processing requirements.

Remote sensing data sets such as satellite imagery or aerial photography will require varying levels of processing depending on how they are received and applied. Ideally, all spatial data sets will be acquired with existing geo-referencing and may only require geographic transformations to be compatible with NPS systems. Varying degrees of spatial and spectral processing may be necessary to adequately address stated inventory and monitoring objectives. The individual protocols outline the necessary processing steps.

Chapter Credits

This chapter was adapted from material prepared by Geoff Sanders (NCRN).

6. Data Quality Assurance

"Data need to meet national-level quality standards and need to be accessible to be used for wise and defensible decision-making at all levels. Data need to be able to be shared and aggregated with data from other parks and from adjacent lands to support landscape-level and national planning and decision-making." (Abigail B. Miller, National Park Service, 2001)

Importance of Quality Assurance/Quality Control (QA/QC)
Ecological data and related information resulting from GRYN resource inventory and monitoring efforts are a valuable resource worthy of preservation only if those data may be used with confidence. Analyses performed to detect trends or patterns in ecosystem processes and the condition of natural resources require data of documented quality. Data of inadequate quality can result in loss of sensitivity to subtle changes and incorrect interpretations and conclusions, and the potential for problems with data quality increases dramatically with the size and complexity of the data set (Chapal & Edwards 1994). Therefore, one important data management goal is to ensure that I&M projects produce data that meet quality requirements specified in monitoring protocols.

Documented methods to ensure data quality are critical to the preservation of data integrity. Established protocols for the identification and reduction of errors at all stages in the data lifecycle, including project planning, data collection, data entry, verification and validation, processing, and archiving, should be incorporated into the data management infrastructure and institutionalized.

While a data set containing zero errors is an ideal goal, such perfection is rarely achieved, and the cost of pursuing 100% accuracy may outweigh the benefit. Two parameters should be considered in setting a data quality goal: 1) the percent of entries that are incorrect (frequency of errors) and 2) the magnitude of the error (criticality of errors). For example, a two-digit numeric entry off by one decimal place is more likely to be a significant critical error than a six-digit numeric entry with the sixth digit off by one. In another case, one incorrect digit in a six-digit species number indicates a completely different species and is clearly significant. Error significance, therefore, is dependent on the type of data. The overall data quality goal should be a reasonable and attainable level of quality based on the intended use of the data and the potential consequences of making a wrong decision or interpretation.

6.1 NPS Mandate for Quality

Producers and users must know and document the quality of their data. This is especially important for sharing data and is the intent of several government directives. NPS Director's Order #11B: Ensuring Quality of Information Disseminated by the National Park Service was issued in 2002 to comply with these directives to ensure and maximize the quality of information disseminated by Federal agencies. The order defines ***'quality'*** as an encompassing term comprising objectivity, utility, and integrity; therefore 'quality' generally refers to all three of these elements. ***'Objectivity'*** includes two distinct elements: 1) presentation, whether disseminated information is being presented in an accurate, clear, complete, and unbiased manner within a proper context and 2) substance, a focus on ensuring accurate, usable, and reliable information. ***'Utility'*** refers to the usefulness of the

information to its intended users, from the perspectives of both the office and the public. ***'Integrity'*** refers to the security of information, e.g., protection from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification. The order further specifies that information will be developed only from reliable data sources and that it will be accurate, timely, and representative of the most current information available. These standards apply not only to NPS-generated information, but also to information provided by other parties to the NPS if the NPS disseminates or relies upon this information.

High quality data and information are not only mandated by directives and orders, they are vital to the credibility and success of the I&M Program. According to Abby Miller (2001) of the Natural Resource Stewardship and Science Division, "data need to meet national-level quality standards and need to be accessible to be used for wise and defensible decision-making at all levels. Data need to be able to be shared and aggregated with data from other parks and from adjacent lands to support landscape-level and national planning and decision-making."

6.2 Definitions of Quality Assurance and Quality Control

Quality assurance procedures plan for quality in all stages of the data development process, while quality control procedures monitor or evaluate the resulting data products. Palmer (2003) defines quality assurance as "an integrated system of management activities involving planning, implementation, documentation, assessment, reporting, and quality improvement to ensure that a process, item, or service is of the type and quality needed and expected by the consumer." He defines quality control as "the overall system of technical activities that measures the attributes and performance of a process, item, or service against defined standards to verify that they meet the stated requirements established by the customer."

QA/QC mechanisms are designed to prevent data contamination, which occurs when a process or event other than the one of interest affects the value of a variable and introduces two fundamental types of errors into a data set. Errors of commission include those caused by data entry and transcription errors or malfunctioning equipment. They are common, fairly easy to identify, and can be effectively reduced upfront with appropriate QA mechanisms built into the data acquisition process, as well as QC procedures applied after the data have been acquired. Errors of omission often include insufficient documentation of legitimate data values, which could affect the interpretation of those values. These errors may be harder to detect and correct, but many of these errors should be revealed by rigorous QC procedures.

QA/QC procedures applied to ecological data include four activities ranging from simple to sophisticated, inexpensive to costly: 1) defining and enforcing standards for electronic formats, locally defined codes, measurement units, and metadata, 2) checking for unusual or unreasonable patterns in data, 3) checking for comparability of values between data sets, and 4) assessing overall data quality. Much QA/QC work is related to the first activity, which begins with data design and continues through acquisition, entry, metadata development, and archiving. The progression from raw data to verified data to validated data implies increasing confidence in the quality of the data through time (Figure 6.1).

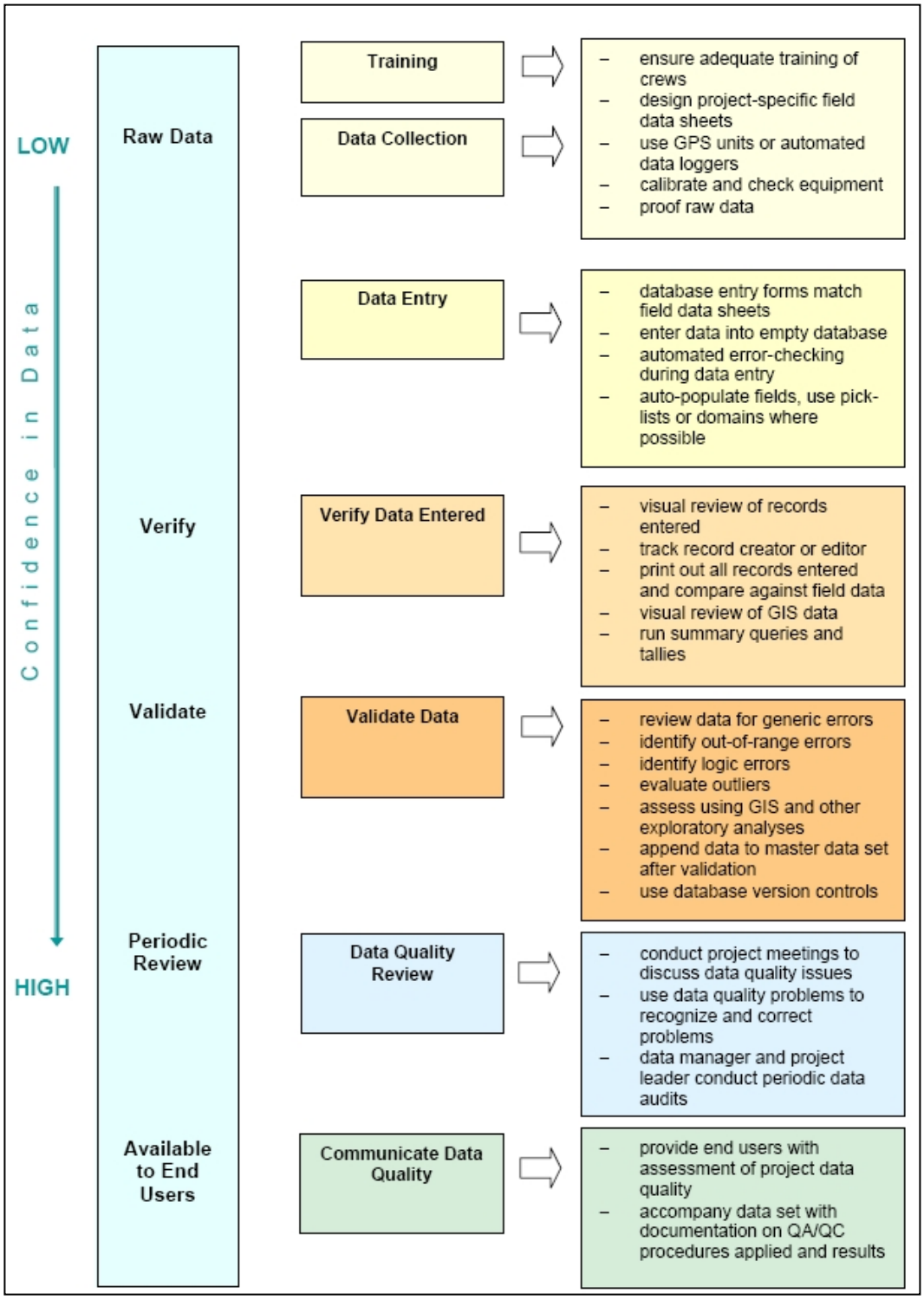


Figure 6.1. QA/QC controls applied at progressive stages of a project (Angel 2004).

6.3 Roles and Responsibilities for Quality Assurance

The importance of planning for quality in data and information before a project begins is critical. Quality assurance methods should be in place at the inception of a project and continue through all project stages to final archiving of the data set. All network employees

from the network coordinator to the data entry technicians, along with the data manager, should take responsibility for, and pride in, achieving a high level of data quality. People are the most important factor in the data quality process, and everyone plays a part in achieving high quality data products. All employees assigned to a project are responsible for the quality of the results generated from his or her task(s). The network coordinator and park managers must recognize the importance of data quality and support all appropriate efforts to achieve the highest possible quality in the data produced by network and park inventory and monitoring projects.

Project leaders must:
1) help prepare quality protocols and convey their importance to technicians and field crews,
2) ensure compliance with the protocols,
3) validate data after the verification process is complete, and
4) review and approve all final reports and information products.

The data manager is responsible for:
1) helping the project leader prepare protocols and SOPs to ensure data quality,
2) making project leaders, technicians, etc., aware of the established procedures and helping enforce adherence to them,
3) evaluating the quality of all data and information against NPS standards before distributing data outside the network, and
4) performing periodic data audits and quality control checks to monitor and improve data quality activities.

Technicians must follow established protocols for data collection, data entry, and verification established in the inventory and monitoring protocol data management SOPs.

6.4 Goals and Objectives for Quality Assurance

The overarching goal in establishing goals, objectives, and criteria for data quality is to ensure that a project produces data of the right type, quality, and quantity to meet the project objectives and the user's needs. Quality criteria should be set at a level proportionate to the project-specific objectives, and these criteria should indicate the level of quality acceptable for the final data product. The EPA (2003) defines data quality objectives as qualitative and quantitative statements that:

- clarify the intended use of the data,
- define the type of data needed to support the decision,
- identify the conditions under which the data are to be collected, and
- specify tolerable limits on the probability of making a decision error due to uncertainty in the data.

The most effective mechanism for ensuring that a project produces data of the right type, quality, and quantity is to provide procedures and guidelines to assist the investigator in accurate data collection, entry, and validation. Therefore, a comprehensive set of SOPs and data-collecting protocols for quality control, namely clear field methodologies, a well-trained field staff, well-organized field forms, and data entry applications with simple built-in validation controls are part of the network's data management approach.

Although specific QA/QC procedures will depend upon the individual vital signs being monitored and must be specified in the protocols for each monitoring vital sign, some general concepts apply to all network projects. The general QA/QC procedures presented in this plan were primarily adapted from the Draft Data Management Protocol (Tessler & Gregson 1997) and the ideas contained in Michener and Brunt (2000). These general guidelines will ensure that all data collected are checked for integrity before being integrated into monitoring program databases. Refer to SOPs and monitoring protocols for specific QA/QC procedures.

6.5 Data Collection

Careful, accurate recording of field observations in the data collection phase of a project will help reduce the incidence of invalid data in the resulting data set. Unlike a typographical error that occurs when a recorded observation is incorrectly transferred from a paper field form to a digital database, an incorrect entry in the field cannot be easily corrected. Therefore, attention to detail during data collection is crucial to overall data quality and will reduce the overall frequency and criticality of errors at subsequent stages in the data lifecycle.

Paper field notebooks and data forms are primary methods for recording ecological data. Paper has advantages in terms of longevity and ease of use, but requires careful handling and storage under some environmental conditions. Data processing options are limited until the data are transferred to digital format. As an alternative to paper, several options for electronic data collection in the field are now available, including handheld computers, automated data collection instruments, and audio recorders (see Chapter 5). Regardless of the collection method, data should be transferred from one form to another only once because each transfer has the potential to introduce additional errors into the data set. One transfer should result in fewer errors, provided that appropriate QA/QC measures are incorporated into the process.

Before the data collection phase of a project, the project leader, with assistance from the data manager provides the protocols/SOPs for data collection and storage. All field sheets and field data recording procedures must be reviewed and approved by the data manager and documented in the protocol SOPs. The project leader ensures that field crew members understand the procedures and closely follow them in the field. The data manager works with the project leader to provide necessary training. Field crew members are responsible for proofing raw data forms in the field, ensuring their readability and legibility, and verifying and explaining any unusual entries. They are expected to understand how to use the data collection forms, know how to take measurements, and follow the protocols.

6.5.1 Suggested Methods to reduce errors during field data collection

Use a formatted, project-specific data sheet as opposed to a field notebook. When electronic data collection devices are not used, data will be recorded on paper data forms. The use of acid-free paper prevents fading and subsequent data loss. Some circumstances require the use of paper and writing implements that can withstand moisture, dust, and other extreme environmental conditions.

Standardized data sheets that identify the parameters to be recorded and display the data for efficient computer entry (i.e., reflect the design of the data entry interface) will help ensure that all relevant information is recorded and subsequent data entry errors are minimized. Data sheets should contain as much basic preprinted project information as possible and sufficient space for recording relevant metadata such as date, collectors, weather conditions, etc. They should clearly specify all required information, using examples where needed to ensure that the proper data are recorded. Project personnel must follow these guidelines:

- All information added to the data sheet must be printed and clearly legible.
- If alterations to the data values are necessary, the original value should be crossed out with a single line and the new value written next to the original entry. Information should never be erased and old information should not be overwritten.
- Upon return from the field, copies of all original data sheets should be made and checked for legibility and completeness (i.e., no data cut off at the edges). The copies of the data sheets will be stored as specified in the protocol SOP, and the original data sheets will be used for data entry.

Use a handheld computer for data collection when applicable. The use of handheld computers minimizes the need for manual data entry from field forms and associated transcription and data entry errors. Specially designed database or computer programs may be required for handheld computers, and the user interface should be customized to meet project requirements. A customized data entry application has the advantage of incorporating on-the-spot QA/QC checks, so this data collection method provides a high level of data quality when combined with point-of-entry data quality checks. These portable units, however, are subject to environmental constraints such as heat, dust, and moisture. When handheld computers are used for data entry in the field, the data must be downloaded daily to avoid potential loss of information. Thus, if a handheld unit fails during data collection, only the current day's data are lost. Batteries should be checked prior to a data collection trip, and they should be charged at the end of every field day. The use of a memory card that will store the data in case of damage to the unit or battery failure is suggested. Finally, in case the unit becomes inoperable in the field, printed data sheets should always accompany field crews on data collection trips.

Use automated data loggers where appropriate. Instruments with their own data acquisition systems may be used to collect some types of data, such as water and air quality data. These devices can be calibrated and programmed to automatically record data and store them for later download directly to a computer, thereby eliminating the possibility for manual data entry errors. Data loggers are an efficient method for recording continuous sensor data, but routine inspections are necessary. Environmental constraints, as well as power (e.g., sufficient battery charge) and maintenance requirements, must be considered in the use of these instruments. Regular downloads are required since physical memory is usually limited, but the elimination of manual data entry reduces potential errors.

Use a handheld tape or digital audio recorder. Another alternative to paper field data forms is a handheld digital audio or micro cassette tape recorder. Recorded observations are subsequently transcribed to paper or directly entered into computer files. As with other technological solutions, there are drawbacks including battery and tape maintenance, low environmental tolerance, and risk of failure. However, if a single data collector is in the field,

audio recorders can provide an easily operated, high quality, efficient method of collecting data.

Consider calibration, maintenance, and minimum timing requirements of field equipment. Accurate field measurements are only possible if field equipment is regularly calibrated and maintained. Where appropriate, consult reference manuals for recommended calibration and maintenance procedures. Once in the field, allow sufficient time for field equipment (such as water quality probes and GPS units) to adjust to the environment so they will record accurate measurements. . Project personnel maintain records of equipment calibration and failures that accompany their field data.

Be organized and keep a log. Organization is the key to good data collection methods. Keeping a log of any decisions made and events that occurred will help clarify information and contribute to an accurate report.

Ensure that field crews receive proper training. Although protocols and SOPs are in place, they cannot guarantee that high quality data will be collected. Prior to routine data collection for a project, conduct training sessions to ensure that field personnel have a clear understanding of data collection procedures described in the SOPs. A training program may also include a process to certify that field staff understand and can perform the specified data collection procedures. The development of a training manual is advised for long-term monitoring data collection efforts and those that will involve a large number of field staff. Palmer and Landis (2002) provide an outline for a training manual and suggestions for planning training sessions.

Perform quantitative assessments of data quality. A quantitative assessment of data quality during data collection activities may be performed to determine if measurement protocols are being followed and quality objectives are being achieved. Repeating a measurement is the primary tool for performing quantitative assessments. Project leaders should periodically review the work of field technicians to ensure that their work does not drift from standards during the course of the field season. Quantitative assessments may be considered if staff and funding are available, and Palmer and Landis (2002) describe several approaches that may be employed.

6.6 Quality Assurance for Data Entry

Data entry is the initial set of operations where raw data from paper field forms are transcribed or typed into a computerized form linked to database tables. Spreadsheets may be used for data entry with approval from the project leader and data manager. When data are gathered or stored digitally in the field (e.g., on a data logger), data entry is the transfer of data (downloading) to a file in an office computer where they can be further manipulated.

Transferring data from field forms into the computer is a straightforward task. Nevertheless, concentration, close attention to detail, and an absence of distractions during data entry are important for reducing errors and maximizing quality. Without proper preparation and some established guidelines, the quality and integrity of the data will be in question. Data entry is best performed by a person who is familiar with the data and ideally takes place as soon as data collection is complete. The goal of data entry is the *transcription* of the data from paper records into the computer with 100% accuracy. However, since transcription errors are

virtually unavoidable during data entry, they will have to be corrected during the data verification process. Observing the data entry guidelines in the next section will help minimize verification work.

The data manager, in conjunction with the project leader, should provide training in the use of the database to all data entry technicians and any other users. The project leader will ensure that data entry technicians understand how to enter data and that they follow the protocols. Data entry technicians are responsible for becoming familiar with the field data forms and differences in handwriting. They must also be familiar with the database software, database structure, and any standard codes for data entry that have been developed. If the program or application used for entering the records is not familiar, the data entry technician should spend some time practicing before actually entering the data. They should know how to open the data entry form, create a new record, and exit the database properly. They must learn how to commit both a 'field' entry and a complete record entry and correct mistakes made while typing.

6.6.1 Suggested Methods to reduce errors during data entry

Enter or download data in a timely manner. All data should be entered or downloaded into the project database as soon as possible, preferably within one week. Do not delay data entry until all the project data have been collected. Downloaded data should be periodically stored on CD or other appropriate media and stored according to the protocol data management SOP.

Design efficient data entry forms and methods. A full-screen data entry form that mimics the field data forms can effectively reduce manual data entry errors due to the 1:1 correspondence of the attributes. A strategy to distinguish between validated data and newly entered data should be adopted. Data may be entered into an empty, fresh database table to avoid contaminating existing data and the new data appended to the master data only after formal verification, validation, and documentation. An alternate method is to include validation attributes that indicate which data have been checked and validated by the project leader in the database. Regardless of the strategy chosen, the process for validation must be clearly documented in the protocol data management SOP and built into the database design.

Build automated error checking features into the database. The most robust QA/QC measures for data entry will be built into the database design to perform automatic validation checks of data. Data entry forms reduce transcription errors through auto-filled fields, range limits, pick lists, and spelling checks; provide controlled access to the database (i.e., forms are set for data entry only, which prevents accidental deletion or alteration of existing data); and control the sequence of data entry (i.e., certain fields require an entry before more information can be entered). Controls warn the operator when errors are made and provide the opportunity for correction before the data are committed to a file.

- *Auto-filled fields.* Whenever possible, the data in a field should be auto-filled by the computer. For example, if a location ID is comprised of a park code, project code, and a unique number, those elements are automatically inserted into the location ID field, ensuring that the record always contains a unique identifier.

- *Range limits.* Where the appropriate values for a particular field span a finite range, the data entry program can check the entered value against the specified minimum and maximum values for that parameter. When a value is outside the accepted range, a warning message appears and asks the user to reenter a valid value. For some fields, values outside a specified 'normal' range may be acceptable. In this case, the warning message asks the user to verify the entry before continuing.
- *Pick lists.* The data entry application may also use pop-up pick lists for standardized text items where spelling errors can occur. For example, rather than typing in a species code or name (where a misspelling generates a new species in the database), the code or name is selected from a list of valid species codes or scientific names and automatically entered into the species field. A pick list may also be used when only certain entries are acceptable. Lists are not appropriate for all written fields but should be used when appropriate.
- *Unique constraints.* Duplicate and incorrect data entry can often be caught with the application of unique constraints on data entry fields. These constraints are particularly useful when importing data from other applications.

Provide a clean, organized work environment. Desktop space near the computer should be free of clutter and distractions that could cause the technician to lose their place. There should be enough space for two stacks of paper documents, one from which data are being entered and one from which data have been entered. A pad or notebook and some fine-point colored markers should also be available for making notes. The need for a clean workspace also applies to the verification and validation phases.

Have two data entry technicians available for data entry. Although not required, when one technician reads the data from the field data forms and another enters them into the computer, the work is often faster and results in a lower error rate. If only one person is available, he/she should work at a slower pace to avoid errors. Like many monotonous tasks, data entry can be done in a personal rhythm that eases the work for some people. Frequent rest periods (approximately every 20 minutes) also helps reduce data entry errors.

Perform initial and interval testing of data quality. To help ensure consistent, useful data is collected for a given monitoring objective it is important to test the data collection procedures and quality control methods soon after field work begins. A mandatory trial period (from one day to two weeks) follows thorough training for personnel involved in data collection. Data from the trial is measured against quality requirements. If the data meets the overall and protocol-specific requirements, then data collection will continue. If the data quality does not meet requirements, then personnel receive additional training and the trial is repeated followed by another quality test. If the data quality does not improve following the second trial, other aspects of the protocol will be examined for factors contributing to the difficulty of meeting data quality requirements.

6.7 Verification and Validation Procedures

Data quality is appraised by applying verification and validation procedures as part of the quality control process. These procedures are more successful when preceded by effective quality assurance practices. Performing both verification and validation of data is emphasized because verified data are not always valid data. ***Data verification*** checks that

the digitized data match the source data, whereas ***data validation*** checks that the data make sense. It is essential that all data are validated as truthful and do not misrepresent the circumstances and limitations of their collection. Failure to follow SOPs for data entry, validation, and verification will render a data set suspect. It is important to remember that only the data entry and verification stages can be done by someone who is not familiar with the kinds of errors sought during validation; validation requires in-depth knowledge about the data.

The project leader and data manager establish SOPs for verification and validation. The project leader will ensure that the SOPs are followed. The project leader or designee will validate the data after verification is complete. He/she is also responsible for reviewing all data products and reports before they are released outside the network. The data and project leaders will evaluate the results of verification and validation and determine any procedural or data form revisions that may be indicated by the results. Technicians will follow the SOPs for verification of data, make required changes, and document those changes.

Manual effort is generally required to get data into electronic format. Any typographical errors made will accumulate in the permanent database unless the data are verified and the errors detected. By implementing data verification procedures, these errors can be reduced, if not eliminated. Data verification immediately follows data entry and involves checking the accuracy of the computerized records against the original source, usually hard copy field records, and identifying and correcting any errors. When the computerized data are verified as accurately reflecting the original field data, the paper forms can be archived.

6.7.1 Suggested Methods for Data Verification

Data verification is most effective when performed immediately after data entry to check the accuracy of the computerized records against the original source, usually hard copy field records, and to identify and correct any errors. Once the computerized data is verified as accurately reflecting the original field data, the original paper forms are archived and further data processing is done on the computer. Each of the following methods has a direct correlation between effectiveness and effort. The methods to identify and eliminate the most errors can be very time consuming while the fastest (cheapest) methods will result in fewer error detections and corrections.

Visually review values and records at the time of data entry. The data entry technician verifies each record after it is input. The values recorded in the database are compared with the original values from the hard copy and any errors are corrected immediately. This method is the least complicated since no additional personnel or software is required. The reliability of this method depends wholly on the person keying data and is generally the least reliable of the data verification methods.

Visually review after data entry. All records are printed upon the completion of data entry. The values on the printout are compared with the original values from the hard copy. Errors are marked and corrected in a timely manner. When one technician performs this review, the reliability increases if someone other than the person keying data performs the review. As an alternative, two technicians may perform this review. One technician reads the original data sheets (the reader) and the second the same data on the printout (the checker).

Perform duplicate data entry. The data entry technician completes all data entry, as normal. Random records are selected (every **n**th record) and entered into an empty replica of the permanent database, preferably by someone other than the person keying the permanent data. A query that automatically compares the duplicate records from the two data sets and reports on any mismatches of data is run. Any disparities are manually reviewed and corrected if necessary. This method involves the overhead of retyping the selected records, as well as the creation of a comparison query (which requires additional effort, but is not time-consuming). This method becomes increasingly successful as the value of **n** decreases. This method is frequently used by professional data entry services.

Generate simple summary statistics. In addition to the above methods, summary information using the entered data can be calculated with statistical software. This is important because even when care is taken up to this point, a duplicate or omitted entry may have been overlooked. For example, the number of known constant elements, such as the number of sampling sites, plots per site, or dates per sample, can be viewed. The same question can be posed in different ways; differences in the answer provide clues to errors. The more checks devised to test the completeness of the data, the greater the confidence that the data are completely verified.

To minimize transcription errors during manual data entry, the network verifies 100% of records to their original source by NPS staff. In addition, 10% of records are reviewed a second time by the project leader, and the results of that comparison are reported with the data. If errors are found in the project leader's review, then the entire data set is verified again.

6.7.2 Data Validation

Data values correctly transcribed from the original field forms into a database are not automatically assumed to be accurate or logical. For example, a pH of 25.0 for a stream is suspect in the database and on the field form. The process of reviewing data for range and logic errors is *validation*. It can be done during data verification *only* if the operator has comprehensive knowledge about the subject matter to which the data relates. More often, validation is a separate operation carried out *after* verification by a natural resource specialist to identify generic and specific errors in particular data types. Validation efforts are often facilitated by database entry controls that inform the user when a value entered is out of a previously defined expected or normal range of continuous values or list of discrete values. For resource specialists not familiar with the database, the data manager can produce hardcopy summary reports and lists for review during the validation phase. Corrections or deletions of logical or range errors in a data set require notations in the original paper field records about how and why the data were changed. Modifications of the field data should be clear and concise while preserving the original data entries or notes (i.e., no erasing!). Validation efforts should also include a check for the completeness of a data set since field sheets or other sources of data could easily be overlooked.

General step-by-step instructions are not possible for data validation because each data set has unique measurement ranges, sampling precision, and accuracy. Nonetheless, validation is a critically important step in the certification of the data. Invalid data commonly consist of slightly misspelled species names or site codes, the wrong date, or out-of-range

errors in parameters with well defined limits (e.g., elevation). But more interesting and often puzzling errors are detected as unreasonable metrics (e.g., stream temperature of 70°C) or impossible associations (e.g., a tree 2 feet in diameter and only 3 feet high). These types of erroneous data are called *logic errors* because using them produces illogical (and incorrect) results. The discovery of logic errors has direct, positive consequences for data quality and provides important feedback to the methods and data forms used in the field. Histograms, line plots, and basic statistics can reveal possible logic and range errors.

6.7.3 Suggested Methods for Data Validation

The following general methods may be used to validate data. Specific procedures for data validation depend upon the vital sign being monitored and will be included in the monitoring protocols.

Data entry application programming. Certain components of data validation are built into data entry forms. The simplest validation during data entry is range checking, such as ensuring that a user attempting to enter a pH of 20.0 gets a warning and the opportunity to enter a correct value between 1.0 and 14.0 (or better yet, within a narrow range appropriate to the study area). Not all fields, however, have appropriate ranges that are known in advance, so knowledge of what are reasonable data and a separate, interactive validation stage are important.

Edwards (2000) suggests the use of 'illegal data' filters, which check a specified list of variable value constraints on the master data set (or on an update to be added to the master) and create an output data set. This output data set includes an entry for each violation, along with identifying information and an explanation of the violation. They illustrate the structure of such a program, written in the SAS® programming language.

A caveat should be interjected regarding the operative word 'illegal'. Even though a value above or below a given threshold has never before been observed and the possibility that it could occur seems impossible, such an observation is not always an illegal data point. Edwards (2000) points out that one of the most famous data QA/QC blunders to date occurred when NASA's computer programs deleted satellite observations of ozone concentrations that were below a specified level, seriously delaying the discovery of the ozone hole over the South Pole.

Outlier Detection. According to Edwards (2000), "the term outlier is not (and should not be) formally defined. An outlier is simply an unusually extreme value for a variable, given the statistical model in use." Any data set will undoubtedly contain some extreme values, so the meaning of 'unusually extreme' is subjective. The challenge in detecting outliers is in deciding how unusual a value must be before it can (with confidence) be considered 'unusually' unusual.

Data quality assurance procedures should not aim to **eliminate** outliers. Extreme values naturally occur in many ecological phenomena; eliminating these values simply because they are extreme is equivalent to pretending the phenomenon is 'well-behaved' when it is not. Eliminating data contamination is a better way to explain this quality assurance goal. If contamination is not detected during data collection, it is usually only be detected later if an outlying data value results. When an outlier is detected, attempts should be made to

determine if some contamination is responsible and to resolve the issue by making appropriate changes and documenting the changes.

Database, graphic, and statistical tools can be used for ad-hoc queries and displays of the data to detect outliers. Some of these outlying values may appear unusual but prove to be valid after confirmation. Noting correct but unusual values in documentation of the data set saves other users from checking the same values.

Other exploratory data analyses. (Palmer and Landis 2002) suggest that in some cases, calculations for assessments of precision, bias, representativeness, completeness, and comparability may be applicable and that for certain types of measurements, evaluation of a detection limit may also be warranted (the authors provide examples of procedures that may be applicable). Normal probability plots, Grubb's test, and simple and multiple linear regression techniques may also be used (Edwards, 2000; the author provides SAS® and Splus® code for constructing normal probability plots and examples of output showing normal and non-normal distributions).

Examples of Validation Errors (Tessler & Gregson 1997)

Wrong Date. A simple typo during data entry creates a logical set of data for a day, month, or year in which samples were never taken. This can become puzzling if the data are sorted by date – thus moving the entry away from its true neighbors. If sorting creates the appearance of missing data where a record should have been, the apparently appropriate corrective action might actually *create* duplicate records in the file rather than fix the ones that were wrong – leaving the original problem unresolved. Even when left in the original order, however, date errors may go undetected because checkers can sometimes see what the readers say – especially when the month and day are the items of focus and an incorrect year digit is not examined. A summary analysis counting the total records in the data set will also be correct. A check of the number of dates or samples per year will often detect an erroneous year by revealing too many samples or a year that does not belong, whereas the rest of the data records reveal where the correction is needed. Identifying site code errors, etc. is a similar process for incorrect values not identified during verification.

Cryptic Duplicates. An example is a contracted vegetation monitoring program. The contractor supplies a field sheet containing species codes, names, and counts in each sample which are to be entered into the computer. Cryptic duplicates occur in the data files when a single sample contains two entries for the same species – the contractor didn't realize he already had a line for that species when doing the counts and added another line later in the table. The data verification process correctly confirms the separate entries but does not recognize that they should be pooled for that sample. Summary counts of the number of species for that sample also show the same number as lines of original data – apparently correct. However, using a *count distinct* query, a count of the number of unique species in the sample is one less than the line count. Returning to the original data form and comparing each line with the others for that sample eventually reveals the error of duplication, and the data file is corrected by pooling the abundance values into the first record of that species and then deleting the second. The original printed data table (the original form) is then also corrected. Here, two different methods of making counts of the same item (species per sample) were used and compared to find the discrepancy. In

Microsoft Access, cryptic duplicates can be eliminated with the proper use of unique primary keys.

Wild Temperatures. Temperatures can show wild variations and yet be completely verifiable and valid. For example, some older data (or the occasional spurious recent record) may have been taken in Fahrenheit rather than Celsius. The difference in the recorded number(s) is large. This is actually a protocol problem and not a data question, but if quality control procedures during data collection were lax, these types of errors are often found only during data validation or (more annoyingly) data analysis. Routinely producing a box-plot or histogram of numerical data reveals drastic outliers, and when the original data forms are consulted, true outliers vs. errors in measurement scale or units become apparent, as does the correction for the files (i.e., convert the measurement to the appropriate units).

Foxes that Change Sexes or Get Younger. As an example, let's use vertebrate monitoring and capture and tagging of foxes each year. The program manager discovered that some of these re-captured measured foxes were changing sex or getting younger. Foxes don't change sexes or get younger. Some serious detective work revealed that inexperienced observers were not able to accurately gauge the age and sex of the foxes. Queries can be run to find these discrepancies. Unfortunately, sometimes these discrepancies cannot be fixed by editing the data files and may result in data of poor quality which may not be usable for their original purpose.

Trees that shrink. The vegetation monitoring program in Shenandoah National Park includes remeasuring trees in permanent plots every five years. In one survey, the project manager discovered that some of the remeasured trees were getting smaller—recent DBHs (diameters at breast height) were less than the original measurements 5 years earlier. Tree trunks of live trees do not get smaller. Some serious detective work revealed that the data were entered accurately (verifiable), but seemingly slight-to-moderate differences in the accuracy and exact methodology by current vs. previous crews. A search-and-compare program was written to parse the data and identify and scale the differences between trees, revealing the extent of the damaged data. Unfortunately, this problem could not be fixed by editing the data files. Rather, it revealed a previous protocol problem that resulted in data of poor quality and data that are useless for the original purpose.

6.8 Version Control

Version control is the process of documenting the temporal integrity of files as they are being changed or updated. Change includes any alteration in the structure or content of the files, and such changes should not be made without the ability to undo mistakes caused by incorrect manipulation of the data. Data progresses through various lifecycle stages, and whenever a set of changes is complete, the user should save the file with a unique name. Version control is simple insurance for maintaining data integrity, and using good version control should be routine for all data handlers.

Prior to any major changes to a file, a copy should be stored with the appropriate version number that allows the tracking of changes over time. With proper controls and communication, versioning ensures that only the most current version is used in any analysis.

The data manager determines the version control method that will be used, and other network personnel are responsible for accurately designating versions for any files upon which they have worked. Staff is encouraged to design and use software tools that assist in file management. For example, databases can be created that include fields to record revision history on the file. Backup routines can be built into the databases that allow for automatic file renaming and archiving. Important program files can be catalogued in a simple index or more formally tracked and archived using professionally developed version control software. Refer to Appendix B for additional information on file management and storage.

Version Control Options

Dates. Using a date provides logical version control. The date is usually formatted as YYYYMMDD or YYMMDD, where DD is optional (depending on the frequency of changes). One drawback to this method is that dates may be hard to read, thus causing confusion for users who may open the wrong version of a file.

Sequential numbers. Versioning of archived data sets is handled by adding a number to the file name, e.g., 001 or V1.0 for the first version. Each additional version is assigned a sequentially higher number. Documenting the date that a file becomes a new version is strongly recommended if this method is used. For example, backup copies of the same database with different raw data are placed in a backup subdirectory with the YYYYMMDD date tagged on the end. Frequent users of the data must be aware of the version control method so they can identify the most recent version.

Version control software. To avoid the work of differentiating multiple versions of documents by appending modifying characters to the file name, version control software is an option. Such software applications track changes made to a document, add comments related to the different document iterations, and retrieve the document at any recorded stage of development. These applications are available in either desktop or online formats.

The following list summarizes some of the issues version control software should address:

- Track documents as they change during the course of the developmental and editorial phases of document/report creation.
- Prevent conflicts between multiple collaborators by prohibiting multiple edits to the same file at the same time.
- Provide the ability to recover previous versions of documents.
- Evaluate the document creation process by tracking who changes a file, when they make the change, and what changes they make.
- Reduce storage requirements by eliminating multiple copies of complete documents.

6.9 Data Quality Review and Communication

QA/QC review is required prior to communicating/disseminating data and information. Only data and information that adhere to NPS quality standards will be released.

Director's Order #11B states that all information (e.g., brochures, research and statistical reports, policy and regulatory information, and general reference information) distributed by

the NPS (including information obtained from sources outside of the NPS) must be accurate, reliable and timely in nature. Therefore, the network must evaluate and identify the types of information it will disseminate that will be subject to the guidelines. Information disseminated to the public must be approved by the appropriate reviewing officials and programs. Documentation of the QA/QC standards used in producing the information and that substantiate the quality of the information must be formally documented Furthermore, mechanisms must be in place for receiving and addressing comments/complaints pertaining to the quality of data.

Data are distributed to the public through the GRYN I&M web page, national web sites such as the Biodiversity Data Store and the Natural Resource/GIS Data Store, and public access databases such as NPSpecies and NatureBib. Any information distributed through any of these mechanisms must undergo internal QA/QC procedures and be approved for release.

Data Quality Review Methods

The network will establish guidelines and protocols to ensure compliance with DO #11B. These protocols will document both internal and external review procedures for data and information disseminated outside the network, as well as a process for processing feedback about data quality.

Edwards (2000) suggests the initiation of quality circles, regular meetings of project leaders, the data manager, and data management personnel for discussing data quality problems and issues. These meetings promote teamwork attitudes while focusing brainpower on data quality issues. Participants become more aware of quality issues and learn to anticipate problems. Moreover, all participants develop a greater appreciation of the importance of their role in data quality and the entire monitoring effort.

Value of Feedback from QA/QC Procedures

Quality assurance procedures may need revision to improve the quality level if random checks reveal an unacceptable level of data quality. Quality checks should not be performed with the sole objective of eliminating errors; the results may also prove useful in improving the overall process. For example, if the month and day are repeatedly reversed in a date field, the data entry technicians may require retraining about the month/day entry order. If retraining is unsuccessful in reducing the error's occurrence, the computer program may need to be rewritten so that month and day are entered separately, field length limits are enforced, or a pick list is created. In this manner, the validation process will serve as a means of improving quality as well as controlling the lack of quality.

Sometimes, modification of field data forms to avoid common mistakes is necessary. With knowledge of validation errors and exploratory data results in hand, the field data forms as the source of the logic errors can be reevaluated. Often minor changes, small annotations, or adding check boxes to a field form remove ambiguity about what to enter on the form. In fact, any time the same type of validation error occurs repeatedly in different data sets, the field form – not the field crew – is usually at fault. Repeated errors found during validation can also mean that protocols or field training are at fault, which can then be recognized and corrected.

Monitoring Conformance to Plans and Standards

Data managers may elect to use periodic data audits and quality control checks as mechanisms to actively participate in the oversight and improvement of data quality activities. Data managers must verify that staff is operating in conformance with the data quality procedures specified in this plan and the protocol specific data management plans. The data manager should track and facilitate the correction of any deficiencies. These quality checks promote a cyclic process of continuous feedback and improvement of the both the data and quality planning process. The cyclic process of quality planning, data collection, data validation and acting on problems is the foundation for the quality assurance philosophies of quality gurus Dr. William Edwards Demming and Dr. Joseph Juran.

Periodic checks by the data manager to see if network staff are adhering to the data quality procedures established in the Data Management Plan and protocols SOPS may include verification of the following:

- Data collection and reporting requirements are being met.
- Data collection and reporting procedures are being followed.
- Verification and validation procedures are being followed.
- Data file structures and maintenance is clear, accurate and according to plan.
- Revision control of program documents and field sheets is adequate.
- Calibration and maintenance procedures are being followed.
- Seasonal and temporary staff have been trained in data management practice.
- Metadata collection and construction for the program is complete.
- Data is being archived and catalogued appropriately for long term storage.

The results of quality assessments should be documented and reported to the project staff and the network coordinator. The project leader and coordinator are responsible for ensuring that non-conformities in data management practices are corrected.

Communicating Data Quality

Data documentation and metadata will be used to notify end users, project leaders, and network management of data quality. A descriptive document for each data set/database will provide information on the specific QA/QC procedures applied and the results of the review. Descriptive documents or formal FGDC-compliant metadata will document quality for spatial and non-spatial data files posted on the Internet.

Every monitoring project requires standard quality control procedures like those listed in table 6.1. In addition to these customary activities, many projects involve unique or specialized quality control measures such as properly calibrating a particular brand and model of equipment used in monitoring.

Table 6.1. Checklist of fundamental quality control activities for a project (Brunt 2000)

Quality control checks	Data Management Phase			
	Design	Acquisition	Metadata	Archive
Check that data sheets represent project design criteria	X			
Check that measurement units are defined on the data sheet	X			
Check that field/attribute names meet NPS & project standards	X			
Check that date, site, and coded values meet NPS & project standards	X			
Check that descriptions of attribute names are provided	X			
Check that data are complete		X		
Check that data entry procedures were followed		X		
Check that data include time, location, and collector(s) full name		X	X	X
Check that measurement data are within the specified range		X		
Check that data values or codes are represented correctly		X		
Check that data are formatted correctly for further use		X	X	X
Check that data table design reflects project design criteria		X	X	X
Check that values for each attribute are used consistently		X	X	X
Check that errors and corrections are recorded		X	X	X
Check that metadata are present			X	X
Check metadata for content (accuracy and completeness)			X	X
Check that data dictionary is present and accurate	X		X	X
Check that measurement units are consistent		X	X	X
Check (again) that data and metadata are complete				X

Chapter Credits

This chapter was adapted from material prepared by Debbie Angel (SODN).

7. Data Documentation

7.1 Purpose of Metadata

Data documentation is information about the content, quality, condition, and other characteristics of a data set so that it may be used and qualified based on its original purpose and limitations. Structured metadata also provide the means to catalog data sets for discovery and distribution on a computer network to a broad range of potential users. In addition to data set documentation, the network will use feature level metadata to document characteristics about each data record associated with a vital sign monitoring protocol. Feature level metadata will include the name, date, source, and version of the monitoring protocol, and FGDC standard attributes for Process Description, Horizontal Positional Accuracy Report, and Process Date. Other feature level metadata characteristics may be added to some data sets.

Many data sets seem to take on lives of their own, with different copies and versions on multiple hard drives, servers, and other storage media. Some data sets remain hidden in digital formats or in seldom used file drawers. Once these data are discovered, a potential user may have little or no information about a data set's quality, content, or process methodologies. Such ambiguity can result in lost productivity when the user must invest time tracking down information, or can render a data set useless because answers to critical questions about the data cannot be found. For these reasons, the planning and implementation of inventory and monitoring projects includes requirements and procedures to provide documentation for all project data. The following mandates and policies exist to guide network staff in developing data documentation strategies.

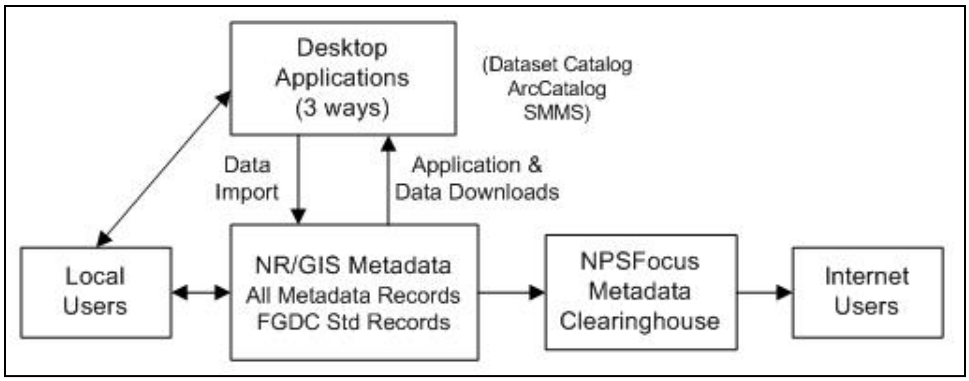
- Executive Order 12906, "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure," signed by President Clinton in 1994, mandates federal agencies to "...document all new geospatial data it collects or produces, either directly or indirectly..." using the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM). In addition, EO 12906 directs agencies to plan for legacy data documentation and provide metadata and data to the public.
- The FGDC Biological Data Profile contains all the elements of the CSDGM and includes additional elements for describing biological data sets. Metadata created in compliance with the Biological Data Profile can be added to the National Biological Information Infrastructure (NBII) Clearinghouse. Although not a requirement, completion of the Biological Data Profile is recommended for biological data sets.
- All GIS data layers must be documented with applicable FGDC and NPS metadata standards. The NPS GIS Committee requires all GIS data layers be described with FGDC standards and the NPS Metadata Profile.
- While there are numerous tools available for developing metadata, the NPS Integrated Metadata System Plan is limited to three recommended desktop applications: Dataset Catalog, ArcCatalog™, and Spatial Metadata Management System (SMMS™).

7.2 NPS Integrated Metadata System Plan and Tools

The NPS Integrated Metadata System Plan represented in figure 7.1 recommends three desktop applications for collecting metadata. The following briefly describes each of these

tools and their potential utility for developing metadata. A fourth tool, the Metadata Parser (mp) is also discussed.

Figure 7.1. NPS Integrated Metadata System Plan



From 12/7/2004 visit to http://science.nature.nps.gov/im/datamgmt/metaplan.htm

Dataset Catalog: Dataset Catalog is a Microsoft® Access database application for cataloging basic metadata on geospatial and biological data sets pertaining to park(s) and networks. It provides a way to inventory, organize, and maintain information about local data set holdings. While Dataset Catalog is not intended to store complete FGDC metadata records, its value for parks and networks is to compile a single list of local data resources along with many of the metadata elements required to meet the mandates of EO 12906. Using the current version of Dataset Catalog (version 2) users can export records as FGDC-format text files for import and further processing using other metadata tools. Dataset Catalog Version 2.1 (in development) will include a function to export records in Extensible Markup Language (XML) that can be imported directly by ArcCatalog™. As the I&M Program recommends, the GRYN catalogs all relevant data sets at I&M parks and the network using Dataset Catalog.

Spatial Metadata Management System: SMMS™ is a commercial software application by Intergraph© for creating and publishing metadata that complies with FGDC requirements. The NPS Integrated Metadata System Plan recommends SMMS™ for FGDC Biological Profile and other geospatial metadata creation. Since the GRYN does not have a license for SMMS™, the network intends to use an NPS-customized ArcCatalog™ metadata extension currently in development.

ArcCatalog™: ArcCatalog™ is a management tool for GIS files contained within the ArcGIS™ Desktop suite of applications.

> "Geographical metadata consists of properties and documentation. Properties are derived from the data source, while documentation is entered by a person. By default, ArcCatalog™ automatically creates and updates metadata, which is stored as well-formed XML data in a file alongside the data or within a geodatabase. Metadata for a folder can also consist of a well-formed HTML file describing its contents." (ESRI®. 2004. ArcCatalog™ Help)

With ArcCatalog™, users can browse, manage, create, and organize tabular and GIS data. In addition, ArcCatalog™ comes with support for several metadata standards that allow one to create, edit, and view information about the data. It includes editors for entering metadata text, a storage schema, and property sheets to view the data. With ArcCatalog™ users can view GIS data holdings, preview geographic information, view and edit metadata, work with tables, and define the schema structure for GIS data layers. Metadata within ArcCatalog™ is stored exclusively as Extensible Markup Language (XML) files. The NPS Integrated Metadata System Plan recommends ArcCatalog™ for gathering GIS-integrated geospatial metadata. The network uses the NPS Metadata ArcCatalog™ Extension developed by NPS Midwest Region GIS Technical Support Center which fixes several ArcGIS™ 8 metadata errors and provides added functionality to properly document NPS data sets. NPS GIS is also developing Biological Profile editing capability and NPS Profile support for the ArcCatalog™ extension.

Metadata Parser: The MetaParser (mp) program is used by the network to validate metadata records by checking the syntax against the FGDC Content Standard for Digital Geospatial Metadata and to generate compliant output files for posting to clearinghouses. It generates a textual report indicating errors in the metadata, primarily in the structure, but also in the values of some of the scalar elements where values are restricted by the standard.

7.3 Metadata Process and Workflow

The general procedures involved with the lifecycle of metadata include identifying which existing and planned data sets require documentation, compiling data characteristics, cataloging, preparing FGDC metadata records, posting the metadata records, and maintaining the records. An overview of these procedures is presented in this section.

Step 1. Identify Data Sets
The scope of data resources requiring metadata for the I&M Program is limited primarily to natural resource themes. Therefore, the network is not concerned with identifying and documenting all data sets for cultural, facility, law enforcement, and other park operations and values. This does not preclude the use of these data by the network for reporting, mapping, etc. Whether natural resource data are acquired as a result of implementing an I&M vital sign monitoring protocol, from existing or future park protocols, or from those of non-NPS sources such as other federal agencies, the value of the data for network parks is determined by professional resource specialists who examine the data and its metadata for relevance to local information needs. For existing data sets determined to be useful and new data sets that are designed to be useful, the next step is to gather information about the data.

Step 2. Compile Metadata Elements
Existing data sets may lack complete metadata and the originator may no longer be available for consultation. In cases where comprehensive documentation is not available, the supporting information relating to a data set is assembled in order to keep track of as much background as possible. Existing data from non-NPS sources may require contacts to request available metadata and/or conduct a metadata interview.

For new projects, metadata development begins up front, with documentation requirements specified in project plans, contracts, and agreements. In addition to written requirements, project leaders, cooperators, and other participants should receive verbal briefings about metadata responsibilities and expectations. Metadata for data sets acquired according to a vital sign protocol will include the protocol name, date, version, and source, along with the protocol's justification statement and specific monitoring objective(s) 'purpose' section of the metadata record. This ensures a link among the complete monitoring protocol documentation, the metadata record, and the data set via the protocol name and version at the record/feature level in the database. The network data manager works principally with project leaders to ensure metadata requirements are met. Complete metadata is required by the network as criteria for completing a project.

Step 3. Create Dataset Catalog Record
The network enters Dataset Catalog records for data sets identified in step one, including both geospatial and non-spatial data sets. Benefits of cataloging brief metadata records include the following:

- provides brief metadata for park and network data holdings in a searchable, centralized location
- helps to organize, identify, and prioritize data sets for which formal FGDC metadata will be developed
- records the status of metadata documentation for a particular data set (i.e., planned, in work, complete)
- supports exporting records for import into the online NR-GIS Metadata System or to continue with additional processing steps based on data type, source, and importance.

Once data sets are cataloged, they can be prioritized for further documentation according to their current and expected level of use or distribution. For example, data sets frequently requested outside the NPS will be fully documented, and those data sets not in use due to vintage or format are considered adequately tracked and documented in Dataset Catalog and the metadata records are available for future processing. All locally published GIS layers will be documented with applicable FGDC and NPS metadata standards.

Step 4. Prepare the Complete Metadata Record
The network plans to use ArcCatalog™ and NPS extensions as the primary tool to create and maintain data set documentation according to FGDC and NPS standards.

In order to correctly process, apply, and interpret raw data contained in the rows and columns of a tabular data files, these must be accompanied by a descriptive document that includes the following information about the project and the data:

- List and description of contents data set content
- Description of the project
- Location of the project study plan and work plan
- Project leader's name and contact information
- Principal investigator's name and contact information
- Data set contact's name and contact information

- Description of the database model (entity-relationship diagram and data dictionary)
- Sensitive data issues, if any
- Description of data verification/validation methods and results
- Additional comments and documentation as needed to inform, as completely as possible, those not involved in the project who will process and apply the data.

Step 5. Make Information Available
Complete metadata and associated data will be submitted to the internet-based NR-GIS Metadata and Data Store as the primary distribution mechanism for sharing information on data holdings with park resource managers, researchers, and others. This system integrates metadata search functions, data downloading, and metadata maintenance. Basic metadata is also available to network parks via reporting functions in Dataset Catalog, including customized queries and reports. Non-sensitive NR-GIS Metadata records are routinely posted to NPS Focus.

Step 6. Maintain metadata records
Like any data set, it takes planning and effort to keep metadata records current. The network intends to schedule annual review of metadata records and to update and synchronize Dataset Catalog and ArcCatalog™ records with the NR-GIS Metadata database.

Chapter Credits
This chapter was adapted from material prepared by Theresa Leibfreid (CUPN).

8. Data Management Support for Analysis and Reporting

Providing meaningful results from data summary and analysis is a cornerstone of the I&M Program and characterizes the network's data management mission to provide useful information for managers and scientists. The Data Analysis and Reporting chapter (VII) of the Vital Signs Monitoring Plan contains the background and overall approach to data analysis and reporting by the network. The associated data management objective is to provide valid data in formats that support scheduled and ad hoc display, query, analysis, summary, and reporting of data to meet local, regional, and national requirements. Routine and scheduled data summary and analysis requirements and procedures are identified in each vital sign monitoring protocol. The following sections discuss data management activities related to using GIS and database application software for data summary and analysis, and to prepare data for analysis using statistical software applications.

8.1 Periodic and Annual Reporting

The network data manager will work with the network ecologist, project leaders, and others involved in data analysis to specify and design or adapt database objects, fields, and values to support the formats and functions necessary for analysis using statistical software applications, e.g., SAS®, R, and S. A list of common vital sign monitoring reports and venues and their purpose, frequency, and intended audience is shown in chapter 7 in the GRYN Vital Signs Monitoring Plan. Some basic summary and reporting functions required by a vital sign monitoring protocol can be developed within Microsoft® Access database applications where data are stored. Examples include descriptive statistics (mean, standard deviation, sample size). The network ecologist and others will use existing and custom data conversion and export functions in Microsoft® Access to prepare data sets for import into other software applications. Spatial analysis and maps will be produced by network and/or park affiliate staff. Ad hoc queries and reports will be handled on a case by case basis due to their dynamic nature.

The network will enhance its web site over time to deliver reports and provide supplemental background data and information. In the future this may include web-based internet map services and functions for user-controlled queries and summaries of network data. The web-site offers easy access to park managers for up-to-date information generated by the network and other related programs.

8.2 Long-term Trends Analysis and Reporting

Most long term data analysis will involve statistical software applications to perform the three primary types of analyses described in Chapter VII of the network Vital Sign Monitoring Plan: parameter estimation, hypothesis testing and model selection. Data formats required by statistical software often involve arrays of binary or discrete values that represent one or more parameters. Data analysts and the network data manager will identify and develop the data conversion routines necessary to generate these formats for analysis.

GIS functions can also contribute to understanding long term status and trends of vital signs and ecosystems. The network will provide geographic data display, summary, and mapping services for scheduled reports and other requests. Methods may be developed for using GIS to visualize time-series data, perform geostatistical, functions, and do spatial network

analysis with hydrography, transportation, and other linear features. Tabular and spatial results can be shared in reports and made available on the network's web site and via internet map services.

9. Data Distribution

9.1 National Park Service Policy on Data Ownership

The National Park Service defines conditions for the ownership and sharing of collections, data, and results from survey and research funded by the United States government. All cooperative and interagency agreements, as well as contracts, should include clear provisions for data ownership and sharing as defined by the National Park Service:

- All data and materials collected or generated using National Park Service personnel and funds become the property of the National Park Service.
- Any important findings from research and educational activities should be promptly submitted for publication. Authorship must accurately reflect the contributions of those involved.
- Investigators must share collections, data, results, and supporting materials with other researchers whenever possible. In exceptional cases, where collections or data are sensitive or fragile, access may be limited.

The Office of Management and Budget (OMB) ensures that grants and cooperative agreements are managed properly. Federal funding must be disbursed in accordance with applicable laws and regulations. OMB circulars establish some degree of standardization government-wide to achieve consistency and uniformity in the development and administration of grants and cooperative agreements. Specifically, OMB Circular A-110 establishes property standards within cooperative agreements with higher institutions and non-profit organizations. Section 36 of Circular A-110, "Intangible Property" describes the following administrative requirements pertinent to data and ownership:

> (a) The recipient (academic institution or non-profit organization receiving federal monies for natural resource inventory and/or monitoring) may copyright any work that is subject to copyright and was developed, or for which ownership was purchased, under an award. The Federal awarding agency(ies) (in this case the National Park Service) reserve a royalty-free, nonexclusive and irrevocable right to reproduce, publish, or otherwise use the work for Federal purposes, and to authorize others to do so.
>
> Section 36 also states:
>
> (c) The Federal Government has the right to:
>
> > (1) obtain, reproduce, publish or otherwise use the data first produced under an award
> >
> > (2) authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes
>
> (d) (1) In addition, in response to a Freedom of Information Act (FOIA) request for research data relating to published research findings produced under an award that were

used by the Federal Government in developing an agency action that has the force and effect of law, the Federal awarding agency shall request, and the recipient shall provide, within a reasonable time, the research data so that they can be made available to the public through the procedures established under the FOIA (5 U.S.C. 552(a)(4)(A)).

(2) The following definitions apply for purposes of paragraph (d) of this section:

(i) Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples)…

(ii) Published is defined as either when:

(A) Research findings are published in a peer-reviewed scientific or technical journal; or

(B) A Federal agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.

(iii) Used by the Federal Government in developing an agency action that has the force and effect of law is defined as when an agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.

9.2 Establishing Data Ownership Guidelines

The network has established guidelines for the ownership of data and other research information. To ensure that proper ownership, format, and development of network products is maintained, all cooperative or interagency work must be conducted as part of a signed collaborative agreement. Every cooperative or interagency agreement or contract involving the network must include OMB Circular A-110 cited under the Reports and Deliverables Section of all agreements and contracts. The following shows appropriate language to use when citing Circular A-110:

> "As the performing organization of this agreement, <*institution or organization name*> shall follow the procedures and policies set forth in Office of Management and Budget (OMB) Circular A-110. (http://www.whitehouse.gov/omb/circulars/a110/a110.html)"

Every cooperative or interagency agreement or contract must include a list of deliverables and products clearly defined within each agreement or contract. Details on formatting and media types that will be required for final submission must be included. Agreements and contracts must list all products expected to result from the project. These include, but are not limited to, field notebooks, photographs (hardcopy and digital), specimens, raw data, and reports.

The following statement must be included in the Reports and Deliverables section of all GRYN agreements and contracts:

> "All reports and deliverables must follow the most recent version of the GRYN Product Specifications."

Investigators should also provide a schedule of deliverables that includes sufficient time for NPS review of draft deliverables before scheduled final submissions.

9.3 Data Distribution

One of the most important goals of the Inventory and Monitoring Program is to integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making.

To accomplish this goal, the network uses procedures to ensure that relevant natural resource data collected by NPS staff, cooperators, researchers and the public are entered, quality-checked, analyzed, documented, cataloged, archived, and made available for management decision-making, research, and education. Providing well-documented data in a timely manner to park managers is important to the success of the program. The network will make certain that:

- Data are easily discoverable and obtainable
- Data that have not yet been subjected to full quality control will not be released to non-NPS requestors, unless necessary in response to a FOIA request
- Distributed data are accompanied by complete metadata that clearly establishes the data as a product of the NPS I&M Program
- Sensitive data are identified and protected from unauthorized access and inappropriate use
- A complete record of data distribution is maintained

To accomplish this, the network uses a variety of distribution methods that allow information collected and developed as part of the program to be widely available to park staff and the public.

9.3.1 Data Distribution Mechanisms

Network staff will coordinate with park staffs to help prevent unnecessary duplication of metadata and data that are posted to internal and external clearinghouses. This will help preserve storage space on the clearinghouse server(s) and prevent confusion among users about duplicate or similar data sets. For distinct but similar data sets the metadata record must include text explaining the differences.

The primary distribution of the network's inventory and monitoring data will occur on the internet. This approach makes data and information available to a broad community of users and does not require a local system to receive and process multiple data requests. As part of the NPS I&M Program, the following web-based applications and repositories (Table 9.1) have been developed to store park natural resource information:

- NatureBib–a master web database housing natural resource bibliographic citations for I&M Program parks (NatureBib web site)

- NPSpecies–a master web-based database to store, manage and disseminate scientific information on the biodiversity of all organisms in all National Park units (NPSpecies web site)

- Biodiversity Data Store–a digital archive of document, GIS data set and non-GIS data set files that document the presence/absence, distribution and/or abundance of any taxa in National Park Service units (Biodiversity Service Center web site)

- Natural Resource and GIS Metadata and Data Store-online repository for metadata and associated data products. NR/GIS Metadata System is a component of the Natural Resource and GIS Data Store (NPS NR-GIS Metadata and Data Store web site prototype).

- GRYN Website–provides detailed information about the I&M network and its activities. Metadata, data sets, and related products such as summary reports for all network-funded inventory and monitoring products will be available via the web site. Park-funded and non-NPS-sourced materials related to the vital signs monitoring program will be linked via the network website where possible.

Table 9.1. Online Database Applications and Related Data Types.

Web Application Name	Data types available at site
NPSpecies	Data on Park Biodiversity (species information)
NatureBib	Park Related Scientific Citations
Biodiversity Data Store	The raw or processed data and products associated with Inventory and Monitoring projects that are stored in NPSpecies.
NR-GIS Metadata and Data Store	Metadata and GIS data sets with no biological component
GRYN Website	Reports and metadata for all Inventory and Monitoring Data produced by the network.

Currently, the NR-GIS Metadata and Data Store and the Biodiversity Data Store are under development. Until procedures and further guidance become available for the use of these two repositories, the network may also disseminate data via the network website. When both Service-wide repositories are completely operational, the network will upload all applicable data and information to each of those sites as needed.

Storing network metadata and data sets in the data stores listed above enables searching via the integrated metadata and image management system and search gateway called NPS Focus. NPSFocus is a Digital Library and Research Station that complies with an international standard protocol (ISO 23950:Z39.50 (http://www.loc.gov/z3950/agency/)) for computer-to-computer information retrieval. This makes NPSFocus a compliant node with the National Spatial Data Infrastructure (NSDI). The system is under development using Blue Angel Enterprise software for metadata management and the LizardTech© Express Server

for image management. One stop searching is currently available for ten NPS and two non-NPS databases in the NPS Focus prototype. NPSFocus is an Intranet version only, and a future public version is planned.

9.4 Data Classification: protected vs. public

All data and associated information resulting from I&M activities must be assessed by network staff to identify sensitive content so that the appropriate level of access can be granted to potential users of the data. This includes, but is not limited to, materials such as reports, metadata, raw and processed spatial and non-spatial data, and maps.

The Federal Geographic Data Committee's Homeland Security Working Group has published an interim version of the "Guidelines for Providing Appropriate Access to Geospatial Data in Response to Security Concerns." The Guidelines include procedures to help identify sensitive information content in geospatial data sets and helps data producers provide appropriate access to the data while protecting sensitive contents.

The Freedom of Information Act, 5 U.S.C. § 552, referred to as FOIA, stipulates that the United States Government, including the National Park Service, must provide access to data and information of interest to the public. FOIA, as amended in 1996 to provide guidance for electronic information distribution, applies to records that are owned or controlled by a federal agency, regardless of whether or not the federal government created the records. FOIA establishes a right for any person to access federal agency records that are not protected from disclosure by exemptions. Under the terms of FOIA, agencies must make non-protected records available for inspection and copying in public reading rooms and/or online. Protected records are provided in response to requests through a specified process. The Department of the Interior's revised FOIA regulations and the Department's Freedom of Information Act Handbook can be accessed at http://www.doi.gov/foia/ for further information. The Northeast Coastal and Barrier Network prepared a summary discussion of FOIA issues related to the Inventory and Monitoring Program and its networks (Appendix C)

In some cases, public access to data can be restricted. Under the NPS Director's Order #66 (draft), and four resource confidentiality laws (the National Parks Omnibus Management Act (16 U.S.C. 5937), the National Historic Preservation Act (16 U.S.C. 470w-3), the Federal Cave Resources Protection Act (16 U.S.C. 4304) and the Archaeological Resources Protection Act (16 U.S.C. 470hh)), the National Park Service is directed to protect information about the nature and location of sensitive park resources. Through these regulations, information that could result in harm to natural resources can be classified as 'protected' or 'sensitive' and withheld from public release (National Parks Omnibus Management Act (NPOMA)).

The following guidance for determining whether information should be protected is suggested in the draft Director's Order #66 (the final guidance may be contained in the Reference Manual 66):

- Has harm, theft, or destruction occurred to a similar resource on federal, state, or private lands?

- Has harm, theft, or destruction occurred to other types of resources of similar commercial value, cultural importance, rarity, or threatened or endangered status on federal, state, or private lands?
- Is information about locations of the park resource in the park specific enough so that the park resource is likely to be found at these locations at predictable times now or in the future?
- Would information about the nature of the park resource that is otherwise not of concern permit determining locations of the resource if the information were available in conjunction with other specific types or classes of information?
- Even where relatively out-dated, is there information that would reveal locations or characteristics of the park resource such that the information could be used to find the park resource as it exists now or is likely to exist in the future?
- Does NPS have the capacity to protect the park resource if the public knows its specific location?

Natural Resource information that is sensitive or protected requires the:

- Identification of potentially sensitive resources
- Compilation of all records relating to those resources
- Determination of what data must not be released to the public
- Management and archival of those records to prevent their unintentional release

Classification of sensitive I&M data will be the responsibility of the GRYN staff, the park superintendents, and investigators working on individual projects. network staff will classify sensitive data on a case by case, project by project, basis. They will work closely with investigators for each project to ensure that potentially sensitive park resources are identified, and that information about these resources is tracked throughout the project.

The network staff is also responsible for communicating all potentially sensitive resources to project leaders and principal investigator(s) working on each project. The investigators, whether network staff or partners, will develop procedures to flag all potentially sensitive resources in all products that come from the project, including documents, maps, images, databases, and metadata. When submitting products and results, investigators should specifically identify all records and other references pertaining to potentially sensitive resources. One method for flagging sensitive information is designing fields in database objects and using data entry procedures for tracking sensitive records. Partners should not release any information in a public forum before consulting with network staff to ensure that the information is not classified as sensitive or protected.

For example, information may be withheld regarding the nature and/or specific locations of the following resources recognized as ‘sensitive’ by the National Park Service. According to NPOMA, if the NPS determines that disclosure of information would be harmful, information may be withheld concerning the nature and specific location of:

- Endangered, threatened, rare or commercially valuable National Park System Resources (species and habitats)
- Mineral or paleontological objects

- Objects of cultural patrimony
- Significant caves

Information already in the public domain can, in general, be released to the public. For example, the media reports in detail the return of condors to the Grand Canyon. If an individual requests site-specific information about where the condors have been seen, general description of geographic areas can be released. However, the locations of specific nest sites cannot be released.

9.5 Access Restrictions on Sensitive Data

Network personnel are responsible for managing access to sensitive data handled by the program. All potentially sensitive park resources will be identified and investigators working on network projects will be informed that:

- All data and associated information must be made available for review by network staff prior to release in any format
- Any information classified as protected should not be released in any format except as approved in advance by the National Park Service

For each project dealing with known or potentially sensitive resources, the network staff provides a complete list of all references to potentially sensitive park resources in each park to the park superintendent for review. Each superintendent then determines which information should be protected.

The network coordinator, project leader, or data manager identifies all potentially sensitive park resources to the principal investigator for each project. Reciprocally, the principal investigators for each project must identify any known references to potentially sensitive park resources.

When preparing to upload information into any network database, the network staff ensures that all protected information is properly identified and marked. The network staff must ensure that all references to protected information are removed or obscured in reports, publications, maps, and other public material.

Network staff will:

- remove any sensitive information from public versions of documents or other media
- isolate sensitive from non-sensitive data
- determine the appropriate measures for withholding sensitive data

The main distribution applications and repositories developed by the I&M Program, (see section 9.3.1) are maintained on both secure and public servers, and all records that are marked ‘sensitive’ during uploading will be available only to appropriate NPS personnel on the secure servers. Procedures for assigning a sensitivity level to specific records when uploading to both the NPSpecies and NatureBib databases are discussed in the GRYN NPSpecies and the NatureBib Data User Manuals (unfinished appendices) as well as at the following websites:

- http://science.nature.nps.gov/im/apps/npspp/index.htm

5. http://www.nature.nps.gov/nrbib/index.htm

Thus, access to data on sensitive park resources can be limited to network staff or project partners. However, limits to how these data are subsequently released must also be clearly defined. It is crucial that the network staff institute quality control and quality assurance measures to ensure that the person doing the uploading of records into the online applications is familiar with the procedures for identifying and entering protected information.

9.6 NPS Inventory and Monitoring Data Availability

According to FOIA (specifically the 1996 amendments), all information routinely requested must be made available to the public via reading rooms and/or the internet. Network project data will be available to the public at one or more internet locations:

- The GRYN web site
- Public servers for the NPSpecies and NatureBib databases
- Public server for the Biodiversity Data Store
- Public server for the NR/GIS Data Store

The network will regularly provide updated information about inventories and monitoring projects, including annual reports and detailed project reports through the network web site. Information on species in the National Parks, including all records generated through the Inventory and Monitoring Program, will be stored and accessible in the NPSpecies database. Bibliographic citations that refer to National Park System natural resources will stored and accessible through the NatureBib database. Documents, maps, and data sets containing resource information from all sources, and their associated metadata, will be accessible through the Biodiversity Data Store and/or NR/GIS Data Store. Each of these databases/repositories offer both a secure server and a public server, and the public can access all information in these databases except those records marked as ‘sensitive.’

Both raw and processed data resulting from the network’s inventory and monitoring projects will be fully documented with FGDC compliant metadata and made available to the public via the network’s website. The metadata for all data sets will be made accessible to the public as soon as they are provided and verified by the principal investigator(s) or project leaders.

Data sets for short-term inventory studies will be provided to the public via the GRYN website not more than two years following the year the data were collected or following publication of the investigator’s results (whichever comes first). Data sets for long-term monitoring studies will be provided to the public when the data is sufficient to support trend analyses reported by the network. The network’s Vital Signs Monitoring Plan and the network web site contain further scheduling details. Before data are posted, a park representative and the investigator or project leader will verify the final data set and metadata if necessary.

GRYN staff will notify investigators prior to making data sets available to the public. This will allow each investigator the opportunity to submit a written request to postpone or further restrict access to the data set by the public. Network staff will review the request to determine whether a data set can or should remain restricted to public access.

Geospatial data records from GIS are a priority electronic records format identified by the US National Archives and Records Administration (NARA) and by partner agencies as part of the Electronic Records Management (ERM) initiative, one of 24 E-Gov initiatives under the President's Management Agenda. A major goal of this initiative is to provide the tools for agencies to access electronic records for as long as required and to transfer permanent electronic records to NARA for preservation and future use by government and citizens. The requirements in this guidance are effective April 15, 2004. The network will incorporate these guidelines into data management operations. http://www.archives.gov/records-mgmt/initiatives/digital-geospatial-data-records.html

9.7 Data Access and Request Procedures

Online Access

The network does not intend to track downloads of data or reports from the network website or other data stores like the Biodiversity Data Store and NR GIS and Metadata Data Store. The network web site will contain a statement about use and appropriate citation of data in resulting publications and a request that users acknowledge the National Park Service Inventory and Monitoring Program.

Offline Requests

For data and information not available online, non-NPS individuals or entities can submit a written request to the network, preferably by email. Requests should include the name of the person or entity, their contact information, the export file format and transfer method, the date delivery is requested, and a description of the content and/or geographic extent of the data. An optional description can be included about the expected use of the data. The purpose of gathering this information in writing is to understand the request for complete and appropriate response by the network, and to document interest in the program and its results by maintaining a list of those who receive network data and products. In many cases the network data manager will follow up by phone with the requesting individual to clarify the information once the written request is received.

9.8 Data Quality Feedback

The GRYN will accept feedback from NPS staff, cooperators and the public on data and information gathered as part of the network's I&M Program. The purpose is to identify, communicate, and resolve errors or concerns about data content and quality. Users can notify network staff by email, phone, or in person. Network staff will investigate and resolve issues and concerns brought to their attention and follow up with the user and other affected individuals to communicate the resolution.

Chapter Credits

This chapter was adapted from material prepared by Sara Stevens (CUPN) and edited by Gary Entsminger (Partner).

10. Data Maintenance, Storage and Archiving

This chapter describes procedures for the long-term management and maintenance of digital data, documents, and objects that result from GRYN projects and activities. The overall goals of these procedures are:

- to preserve information over time
- to ensure that information can be easily obtained, shared, and properly interpreted by a broad range of users.

Effective long-term data maintenance is inseparable from proper data documentation, and an essential part of any archive is accompanying explanatory materials (Olson and McCord 1998). This chapter will refer to, and in some cases elaborate on, metadata standards and data set documentation procedures that are more fully explained in Chapter VII (Data Documentation) of this plan.

10.1 Digital Data Maintenance

In general, digital data maintained over the long term will be one of two types: short-term data sets for which data collection and modification have been completed (i.e. inventory projects), and long-term monitoring data sets for which data acquisition and entry will continue indefinitely.

Technological obsolescence is a significant cause of information loss, and data can quickly become inaccessible to users if stored in out-of-date software programs, on outmoded media, or in legacy data formats. Maintaining digital files involves managing for the dynamic infrastructure of associated hardware, software, file formats, and storage media. Major changes in hardware can be expected every 1-2 years, and in software every 1-5 years (Vogt-O'Connor 2000). As software and hardware evolve, data sets must be consistently migrated to new platforms, or they must be saved in formats that are independent of specific platforms or software (e.g., ASCII delimited files).

Data sets for which data entry or updates are still occurring will be stored in a structured file system for active projects on the GRYN server. A file system for completed projects will hold data and information resources that are no longer expected to change.

10.1.1 Short-term data sets

For finalized short-term data sets created or managed by GRYN, a set of American Standard Code for Information Interchange (ASCII) tab-delimited or comma-delimited text files will be created for each data table comprising the data set. These files will be accompanied by a readme.txt file that explains the contents of each file, file relationships, field definitions, and useful queries in sql format. Because delimited ASCII files can be imported by virtually all known database and spreadsheet software applications, the ASCII files serve as a format independent, cross-platform archive of the native version of the data set. All finalized files will be stored on the GRYN server in the appropriate project's archive folder.

In addition to creating ASCII files, the network will update data sets stored in previous versions of Microsoft® Access, so that no data set is more than two versions behind the

current version of Microsoft® Access used by GRYN. To the extent possible, preexisting functionality of data entry forms and reports will be maintained. However, the priority will be to ensure basic table and relationship integrity.

10.1.2 Long-term monitoring data sets

Long-term monitoring data sets require regular updates and conversion to current database formats. All active or long-term databases will conform to the current NPS and I&M software version standards. Monitoring projects have variable long-term data archiving requirements that include those listed here.

- Original data sets are normally stored in perpetuity as a basis for processing and deriving new values. This can be accomplished both by preserving values in original data fields throughout the life of a database object, and by archiving complete database objects.
- Substantial changes in database design or content due to evolving monitoring protocols may require complete data sets to be archived, with new data stored in different database structures and/or formats. If necessary for comparing new and legacy data sets, a process to relate between the two should be developed as part of the transition.
- Some project methods require preservation of interim data sets (data "milestones") over the long term.

Data sets or subsets destined for long-term archiving will be saved, whenever possible, in their native formats in addition to ASCII delimited text files. Data archiving requirements for ongoing projects will be spelled out in the data management SOPs for each monitoring project.

10.2 Quality control for converted data

All ASCII files created from databases will undergo quality control activities or functions to ensure that the number of records and fields correspond to the source data set, and that conversion has not created errors or data loss. A second reviewer (preferably a program scientist) will evaluate the ASCII files and documentation to verify that tables, fields, and object relationships are adequately explained and documented for users not familiar with the program.

Databases that are converted from one version of database software to an upgraded version will require additional QC, particularly when the database applications are actively used for data entry or analysis. Forms, queries, reports, and data entry will be thoroughly tested during upgrades.

Version control

Documentation of version updates and associated details will be part of the archive metadata record, and revision information and history will be included in tables within the database files. File names of the archived revisions will clearly indicate the revision number or date.

Spatial data

Spatial data sets are maintained in formats that remain compatible with the current version of ArcGIS™. The network will work towards converting existing GIS formats to the personal

geodatabase format. The network will investigate using the Spatial Data Transfer Standard (SDTS) to archive GIS data in a software or platform-independent format. Both uncorrected and corrected GPS data (e.g., Trimble .ssf and .cor files) will be archived in their native format in addition to the corresponding GIS files that are created. Remote sensing data, including satellite images, are stored in their native format and in processed formats for use in image processing and GIS applications.

Digital Still Images
The network expects that most images acquired with handheld cameras will be still images in digital Joint Photographic Experts Group (JPEG) format. Network staff and cooperators are encouraged to use digital camera resolution and quality settings that produce images with adequate quality at a reasonable file size. Project crew members normally collect more images than necessary for documenting a project. The project leader should select and submit only those images necessary to complement the other project data. The data manager will review images for quality and file size and may process the images to meet specifications. Digital images are stored and named according to the instructions in each monitoring protocol.

10.3 Storage and Archiving Procedures for Digital Data

Digital data need to be stored in a repository that ensures both security and ready access to the data in perpetuity. GRYN will maintain a RAID (redundant array of independent disks) server for data storage, combined with a schedule of full, incremental, and differential backups using Network Attached Storage (NAS) devices for on-site backup and magnetic tape for off-site backup.

10.4 Directory structure for individual projects

The organization and naming of folders and files should be concise, descriptive, and clear to users unfamiliar with a specific project. Because the amount and type of material varies with each project, a high level standardized folder structure organizes several of the following common project elements:

- administrative documents such as agreements, correspondence, research permits
- programmatic documents including protocols, procedures, supporting documents
- interim data sets or “milestones”
- data sets distinguished by status: original, in-work, published, and archived
- conceptual or statistical models used for data interpretation
- final reports
- ‘readme’ files - includes an explanation of directory contents, project metadata (including a Dataset Catalog report), and version documentation.

Files are stored in a hierarchical folder structure that provides context to the information system. Folder structures are kept as flat as possible while providing for adequate classification. The network uses the folder and file naming and version tracking conventions in Appendix B.

Once final data and reports have been submitted, approved, and archived, draft and intermediate products should be reviewed and deleted if appropriate.

10.5 Backup procedures for digital data

The risk of data loss exists from a variety of sources including catastrophic events (e.g., fire, flood), user error, hardware failure, software failure or corruption, and security breaches or vandalism. Network staff perform regular backups of data and arrange for off-site storage of backup sets to safeguard data resources.

Data stored on the GRYN server receives weekly normal backups and nightly differential backups using a Network Attached Storage device. A set of four revolving quarterly full backups on magnetic tape are stored off-site to preserve all digital information resources for one full year. All backups are performed and monitored by the network data manager. The data manager tests backup files at random by performing intermittent restore operations. All backups are run with "verify," which compares source files against the backup files and detects any discrepancy in file size or other errors. Magnetic tape media is replaced every two years to prevent data loss due to deterioration.

Backups of data that reside on the desktop computers of network staff are the responsibility of each staff member. Network staff is reminded regularly to copy working files onto the network server or local backup device. The data manager works with network staff and cooperators to develop and perform regular backups of network data resources.

10.6 Data and network security

Access to files on the network server requires successful completion of all required security awareness and IT training for both I&M staff and network partners. Folders containing completed or protected project data may be designated as read-only for certain staff, as determined by the network coordinator. This helps ensure that changes to files are authorized by appropriate staff and communicated to the data manager when updates to associated documentation or other activities must accompany the change. It also deters inadvertent copying, moving, or deleting of files and folders.

10.7 Storage and Archiving Procedures For Documents and Physical Objects

This section applies to documents such as final reports prepared by staff or contractors, program administrative documents, contracts and agreements, memoranda of agreement, and other documents related to GRYN administration, activities, and projects. This section also applies to physical items such as natural history specimens, photographs, and audio tapes. In most instances these documents and objects are essential companions to the digital data described in the previous sections.

Direction for managing many of these materials (as well as digital materials) is provided in NPS Director's Order 19: Records Management (2001) and its appendix, NPS Records Disposition Schedule (NPS-19 Appendix B, revised 5-2003). NPS-19 states that all records of natural and cultural resources and their management are considered mission-critical records, that is, necessary for fulfillment of the NPS mission. NPS-19 further states:

> Mission critical records are permanent records that will eventually become archival records. They should receive the highest priority in records management activities and resources and should receive archival care as soon as practical in the life of the record.

Section N of NPS-19 Appendix B, which provides guidelines on natural resource-related records (including, specifically, the results of Inventory and Monitoring programs), indicates that all natural resource records are considered "permanent," that is, are to be transferred to the National Archives when 30 years old. It also indicates that non-archival copies of natural resource-related materials are "...potentially important for the ongoing management of NPS resources" and should not, in any instance, be destroyed.

10.7.1 Documents

All paper documents managed or produced by the GRYN will be housed in one of three locations:

1. GRYN central files, Bozeman, MT.
Central files contain project files, administrative documents, and non-record copies of documents that are archived at an off-site facility (see item 2, below). Examples include: meeting minutes, correspondence, memoranda of understanding, contracts and agreements, research permits, interim and selected final reports produced by the program or under its auspices. GRYN will use acid-free paper and folders for all permanent records in the central files. In addition to maintaining these paper records, GRYN will maintain electronic versions, when possible, on the GRYN server. The central files are maintained by the GRYN administrative clerk, under the guidance of the data manager and network coordinator.

2. Western Archeological and Conservation Center (WACC)
[repository subject to change]. WACC provides temperature and humidity-controlled facilities, a professional archival staff, and meets all museum standards set by NPS. This repository will be used for original documents and associated materials produced by the network (e.g., photographs, field notes, permits) that are a high priority to maintain under archival conditions. Examples include: original inventory reports and accompanying slides and maps; original vegetation mapping reports; GRYN Phase 1, 2 and 3 reports. Copies of these reports will be maintained in the GRYN central files, and all will have an electronic equivalent (e.g., PDF) for distribution or reproduction.

For all materials submitted to WACC, GRYN will provide essential cataloging information such as the scope of content, project purpose, and range of years, to facilitate ANCS+ record creation and accession. GRYN will also ensure that materials are presented using archival-quality materials (acid-free paper and folders, polypropylene or polyethylene slide pages).

Many GRYN reports and documents encompass data from multiple parks, which makes it challenging to accession archival copies into a specific network park museum. In these instances WACC will prepare associated ANCS+ records that reference all parks included in a report or document, and will prepare finding aids to help potential users locate the materials.

3. Network parks' central files and/or museums.
High-quality copies of park-related documents resulting from GRYN projects, along with electronic versions, will be provided to park resource management staff. Parks may choose to accession these materials into their museums, incorporate them into their central files, or

house them in their resource management library. GRYN will not directly manage documents at the park level but will work with park resource managers and curators to bridge network and park level document management requirements.

10.7.2 Specimens

The network assists with planning for monitoring projects that includes budgeting for specimen preparation and records processing and cataloging. Specimens collected under the auspices of GRYN will be provided to the network park in which they were collected for curation, or to a repository approved by a park (where the specimens are considered on loan). GRYN will provide park curators with necessary data for cataloging each specimen. This data will be in comma-delimited format (.csv) for automated uploading into ANCS+. Data provided to non-NPS curators will be in an appropriate format for each institution.

10.7.3 Photographs

Due to concern among some archivists that the accelerating rate of technological change threatens documentary heritage in the information age (Cox 2000), GRYN accepts and processes 35mm slides (preferably Kodachrome or Ektachrome), which have a proven long-term stability (Wilhelm and Brower 1993), and 4x6 color prints. Original photographs are a high priority for placing in archival storage conditions.

Slides are labeled using indelible pigment ink, or using laser-printed archival-quality slide labels. Slide labels will include: a unique ID, project name, photographer name, photo date, a brief identification of contents (e.g., species name, plot ID), and geographic location (UTM coordinates or a description). All slides are stored in polypropylene slide sleeves at the GRYN office until transferred to WACC. In addition, all slides are scanned and saved digitally as Tagged Image File Format (TIFF) files, and these are used as the primary means of distributing and reproducing the images.

Photographs are stored in individual polypropylene sleeves and within archival boxes. Each photo is labeled on the back, using archival-quality labels that are either laser-printed or hand-labeled with the same information elements required for slides. If a contractor is submitting photographs, corresponding TIFF files must also be submitted.

Every image, regardless of format, has an entry into the GRYN Photo Database where attributes such as electronic file name, keywords, project, photo description, photographer, date, and location are catalogued. All digital image files and the associated photo database are stored on the GRYN server.

10.8 Role of curators in storage and archiving procedures

Curators for parks within GRYN and from other NPS collection facilities are a valuable source of expertise, advice, and guidance on archiving and curatorial issues, and they have a role in almost every project undertaken by the network. Project leaders should involve park curators when projects are in the planning stage to ensure that all aspects of specimen curation and document archiving are considered, and that any associated expenses are included in project budgets.

Chapter Credits
This chapter was adapted from material prepared by Margaret Beer (NCPN).

Change Procedures and Revision History Log

This following table lists all edits and amendments to this document since the original publication date. Information entered in the log must be complete and concise. Users of this document should promptly notify the GRYN data manager and/or program coordinator about recommended and required changes. The GRYN data manager must review and incorporate all changes, complete the revision history log, and change the date and version number on the title page and in the footer of the document file.

Write down edits and versioning in the Revision History Log table for the GRYN Data and Information Management Plan. Version numbers increase incrementally by hundredths (e.g. version 1.01, version 1.02, …etc) for minor changes. Major revisions should be designated with the next whole number (e.g., version 2.0, 3.0, 4.0 …).

Post new versions on the GRYN web site and communicate the update to all known individuals with a previous version of the Plan.

Revision History Log:

Original Publication Name, Version and Date: Data and Information Management Plan - Greater Yellowstone Inventory and Monitoring Network, Version 1.0, September 30, 2005				
Date and number of Previous Version	Date of Revision	Author(s) of Revision (with title and affiliation)	Location in Document and Concise Description of Revision	Reason for Change
Add rows as needed for each change or set of changes tied to an updated version number				

References

Angel, D. 2004. Sonoran Desert Network Data Management Plan, National Park Service.

Boetsch, J.R., B. Christoe, and R.E. Holmes. 2004. Draft data management plan for the North Coast and Cascades Network Inventory and Monitoring Program. National Park Service. Port Angeles, WA.

Brunt, J. W. 2000. Data management principles, implementation and administration. Pp. 25-47 *in* Michener, William K., and James W. Brunt (eds), 2000, Ecological Data: Design, Management and Processing, Methods in Ecology Series, Malden, MA: Blackwell Science Ltd.

Chapal, S. E., and Edwards, D. 1994. Automated smoothing techniques for visualization and quality control of long-term environmental data. Pp. 141-158 in *Environmental Information Management and Analysis: Ecosystem to Global Scales* edited by W. K. Michener, J.W. Brunt, and S.G. Stafford. London: Taylor & Francis Ltd.

Cox, R. J. 2000. Searching for authority: archivists and electronic records in the New World at the fin-de-siècle. First Monday 5:1. <http://firstmonday.org/issues/issue5_1/cox/index.html>. Accessed 2005 September 12.

Edwards, D. 2000. Data quality assurance. Pp. 70-91 in *Ecological Data: Design, Management and Processing, Methods in Ecology Series* edited by W. K. Michener and J.W. Brunt. Malden, MA: Blackwell Science Ltd.

Environmental Systems Research Institute© ArcCatalog™ (Build 800). 2002. Help Menu.

Federal Geographic Data Committee. USGS, 590 National Center, Reston, VA 20192 http://www.fgdc.gov/metadata/metadata.html. Accessed 2005 September 12.

Gregson, J. 2000. Natural Resources Data and Information Systems Handbook Outline. USDI National Park Service Inventory and Monitoring Program. Unpublished.

Gregson, J. 2002. Draft Natural Resource Data and Information Systems Management Handbook Chapter 2. USDI National Park Service Inventory and Monitoring Program. <http://www1.nrintra.nps.gov/im/datamgmt/nrdismhb.htm>. Accessed 2004 August 29.

International Organization for Standardization. < http://www.iso.org>. Accessed 2005 September 12.

McGinnis, J. and Wickwire, D. 1998 Draft. Spatial Data Management, A Shared Perspective. USDI Bureau of Land Management and USDA Forest Service. <http://ims.reo.gov/website/swop/html/swop_slide_show/swop_sdm_whitepaper/sdm.htm>. Accessed 2005 September 29.

Miller, A. B. 2001. Managing Data to Bridge Boundaries. Pp. 316-320 in *Crossing Boundaries in Park Management: Proceedings of the 11th Conference on Research and Resource Management in Parks and on Public Lands*, edited by David Harmon. Hancock, Michigan: The George Wright Society

Office of Management and Budget. Circular A-110 <http://www.whitehouse.gov/omb/circulars/a110/a110.html>. Accessed 2005 September 12.

Olson, R.J. and McCord, R.A. 1998. Data Archival. pp 53-57 in *Data and information management in the ecological sciences: a resource guide* Michener, W.K., J.H. Porter, and S.G. Stafford. LTER Network Office, University of New Mexico. Albuquerque, NM.

Palmer, C. J. 2003. Approaches to Quality Assurance and Information Management for Regional Ecological Monitoring Programs. Pp. 211-225 in *Monitoring Ecosystem: Interdisciplinary Approaches for Evaluating Ecoregional Initiatives* edited by D. E. Busch and J.C.Trexler. Washington, DC.: Island Press

Palmer, C. J. and E. B. Landis. 2002 Draft. Lake Mead National Recreational Area Resource Management Division: Quality System Management Plan for Environmental Data Collection Projects.

Powell, D. S. 2000. A "White Paper" for the Washington Office Ecosystem Management Corporate Team & Interregional Ecosystem Management Coordinating Group, United States Department of Agriculture Forest Service, Washington, D.C.

Smith, S. and Truitt, B. 2004. Klamath National Parks Network Data Mining Protocol, Klamath Network, National Park Service.

Tessler, S. and Gregson J. 1997. Draft Data Management Protocol. USDI National Park Service Inventory and Monitoring Program. <http://www1.nrintra.nps.gov/im/dmproto/joe40001.htm>. Accessed 2004 August 21.

National Water Quality Monitoring Council < http://water.usgs.gov/wicp/acwi/monitoring/>. Accessed 2005 September 12.

United States Government Accountability Office. August 2005. DATA MINING - Agencies Have Taken Key Steps to Protect Privacy in Selected Efforts, but Significant Compliance Issues Remain. http://www.gao.gov/new.items/d05866.pdf. Accessed 2005 September 9.

United States Government Accountability Office. May 2004. DATA MINING – Federal Efforts Cover a Wide Range of Uses. http://www.gao.gov/new.items/d04548.pdf. Accessed 2005 September 9.

US Environmental Protection Agency. 2003. Guidance for geospatial data quality assurance project plans. EPA QA/G-5G. EPA/240/R-03/003. Washington, DC: Office of Environmental Information.

US Environmental Protection Agency. STORET. < http://www.epa.gov/storet/>. Accessed 2005 September 12.

USDA Forest Service Forest. 2004. Forest Service Natural Resource Applications Geodatabase Design Tactical Plan. Unpublished.

USDA Forest Service. 2004. Resource Information Management Roles and Responsibilities, Intermountain Region, Unpublished.

USDA. Integrated Taxonomic Information System http://www.itis.usda.gov/. Accessed 2005 September 12.

USGS National Biological Information Infrastructure. < http://www.nbii.gov/>. Accessed 2005 September 12.

USGS National Hydrography Dataset. < http://nhd.usgs.gov/>. Accessed 2005 September 12.

USDI National Park Service. 2000. Greater Yellowstone Network Vertebrate and Vascular Plant Inventory Study Plan. National Park Service, Washington D.C.

USDI National Park Service. 2003. NPS records disposition schedule. NPS-19, Appendix B – Revised, 5-03.

USDI National Park Service. NPS Natural Resource Challenge. < http://www.nature.nps.gov/challenge/>. Accessed 2005 September 12.

USDI National Park Service. The IT Manager's Playbook – Guidelines and Standards for NPS IT Policies. National Park Service, Washington, D.C.

USDI National Park Service. Vital Signs Water Quality Data Management and Archiving. < http://www.nature.nps.gov/water/infoanddata/>. Accessed 2005 September 12.

USDI National Park Service. 2002. Director's Order #11B: Ensuring quality of information disseminated by the National Park Service. National Park Service, Washington, D.C.

USDI National Park Service. 2005. Greater Yellowstone Inventory and Monitoring Network Vital Signs Monitoring Plan. National Park Service, Bozeman, MT.

Vogt-O'Connor, D. 1997. Caring for photographs: general guidelines. National Park Service Conserve O Gram 14:4. NPS Museum Management Program, Washington, DC.

Vogt-O'Connor, D. 2000. Planning digital projects for preservation and access. National Park Service Conserve O Gram 14:4. NPS Museum Management Program, Washington, DC.

Wilhelm, H. and C. Brower. 1993. The permanence and care of color photographs: traditional and digital color prints, color negatives, slides, and motion pictures. Preservation Publishing Company: Grinnel, IA.

Appendix A: Roles, Responsibilities, and Awareness Levels for Data Management

This table is an expanded version of Table 2.2 – Summary of Roles and Responsibilities in the Data and Information Management Plan. This representative list of data stewardship roles, responsibilities, and awareness levels can be applied and extended by anyone involved with the NPS Greater Yellowstone Inventory & Monitoring Network. For example, the listed elements can help to brief and debrief project staff and crew members, and inform training plans, work plans, recruitment efforts, contract specifications, and other data management aspects of network operations.

Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
Project Crew Member	Collect, record, and verify data	Obtain training in data management for the project. Read and follow project protocols, study plans, and relevant NPS guidance. Communicate with Crew Leader, Project Leader, and Data Manager. Record and verify observed or measured data values. Schedule and perform regular data transfer and backup. Review, verify, and correct field data. Assist with data and procedural documentation, especially deviations from the protocol or study plan.	Aware of specific protocol and related subject(s) for the project. Aware of related protocols and projects.
Project Crew Leader	Supervise crew	Obtain training in data management for the project. Ensure crew members receive data management training and briefings. Conduct debriefing sessions; summarize and communicate the results to the project leader. Read and follow all protocol, project, and relevant Network-level guidelines. Communicate with Crew Members, Project Leader, and Data Manager. Ensure data are regularly transferred, backed up, verified, and entered into the appropriate NPS database(s). Assist with data and procedural documentation.	Aware of specific protocol and related subject(s) for the project. Aware of related protocols and projects. Aware of Network activities, NPS I&M Program generally.
Data/GIS Specialist or Technician	Process and manage data	Obtain briefings about projects and related data to understand the geospatial and technical requirements and relevance. Communicate with other participants in the project to the extent necessary to accomplish assigned tasks. Perform assigned level of technical data management and/or GIS activities, including data entry, data conversion, and documentation. Work on overall data quality and stewardship with Project Leaders, Resource Specialists, and the Network Data Manager.	Aware of Network activities, NPS I&M Program generally. Aware of data management relationships between parks and the network.

Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
Information Technology/ Systems Specialist	Provide IT/IS support	Provide and maintain an information systems and technology foundation to support data management. Advise project participants about capabilities of hardware and software resources to support project and program objectives. Work with Database Manager to resolve hardware and software issues relating to database functions and availability.	Aware of NPS I&M Program generally. Aware of database tools and applications used by the I&M program.
Project Leader	Oversee and direct project operations	Ensure Crew Leader receives pertinent training and briefings. Prepare debriefing plans and materials for field crews; participate in selected debriefing sessions. Communicate with Crew Leader, Data Manager, and I&M Network Coordinator. Complete project documentation describing the who, what, where, when, why and how of a project. Develop, document and implement standard procedures for field data collection and data handling. Enact and supervise quality assurance and quality control measures for the project. Supervise and certify all field operations, including staff training, equipment calibration, species identification, and data collection. Supervise or perform data entry, verification and validation. Maintain concise explanatory documentation of all deviations from standard procedures. Ensure documentation of important details of each field data collection period. Maintain hard copies of data forms and send original data forms to archive on a regular basis. Work with program coordinators to identify analysis and reporting mechanisms, and to establish a schedule for regular project milestones such as data collection periods, data processing target dates, and reporting deadlines. Produce regular summary reports and conduct periodic trend analysis of data, store the resulting reports, and make them available to users. Act as the main point of contact concerning data content. **The project leader works closely with the data manager to:** Develop quality assurance and quality control procedures specific to project operations. Identify training needs for staff related to data	Aware of specific protocol and related subject(s) for the project. Aware of related protocols and projects. Aware of Network activities, NPS I&M Program generally. Aware of park resource information management and GIS Plans and capabilities. Aware of other related projects external to NPS.

Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
Project Leader (continued)	Oversee and direct project operations	management philosophy, database software use, quality control procedures, etc. Coordinate changes to the field data forms and the user interface for the project database. Fully document and maintain master data. Identify sensitive information that requires special consideration prior to distribution. Manage the archival process to ensure regular archival of project documentation, original field data, databases, reports and summaries, and other products from the project. Define how project data will be transformed from raw data into meaningful information and create data summary procedures to automate and standardize this process. Identify and prioritize legacy data for conversion; convert priority data sets to a modern format. Increase the interpretability and accessibility of existing natural resource information. **Note:** The Project Leader is often a resource specialist, in which case the associated responsibilities for data authority apply (see resource specialist role). A Project Leader without the required background to act as an authority for the data will consult with and involve the appropriate Resource Specialists.	
Resource Specialist	Understand the project and make decisions about the data	Understand the objectives of the project, the resulting data, and their scientific and management relevance. Guide development of an Information Needs Assessment based on the objectives of the project. Make decisions about data with regard to validity, utility, sensitivity, and availability. Describe, publish, release, and discuss the data and associated information products. **Note:** The Resource Specialist serving as a Project Leader is also responsible for the duties listed with that role.	Aware of specific protocol and related subject(s) for the project. Aware of related protocols and projects. Aware of Network activities, NPS I&M Program generally. Aware of park resource information management and GIS Plans and capabilities. Aware of other related projects external to NPS.

Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
GIS Manager	Support park management objectives with GIS and resource information management	Coordinate and integrate local GIS and resource information management with Network, Regional, and National standards and guidelines. **The GIS specialists will work in collaboration with project leaders to:** Determine the GIS data and analysis needs for the project. Develop procedures for field collection of spatial data including the use of GPS and other spatial data collection techniques. Display, analyze, and create maps from spatial data to meet project objectives. Properly document data in compliance with spatial metadata standards. **GIS specialists will also work directly with data managers to:** Design databases and other applications for the network. Create relationships between GIS and non-spatial data and create database and GIS applications to facilitate the integration and analysis of both spatial and non-spatial data. Establish and implement procedures to protect sensitive spatial data according to project needs. Develop and maintain an infrastructure for metadata creation and maintenance. Ensure that project metadata are created and comply with national and agency standards.	Aware of Network activities, NPS I&M Program generally. Aware of data management relationships between parks and the network. Aware of associations between park resource issues and park/network I&M objectives.
Network Data Manager	Ensure inventory and monitoring data are organized, useful, compliant, safe, and available	Assist in developing and implementing procedures to ensure that I&M data collected by NPS staff, cooperators, researchers and others are entered, quality-checked, analyzed, reported, archived, documented, cataloged, and made available to others for management decision-making, research, and education. Provide guidance and support, to the extent possible, to extend Network standards and procedures to studies and data funded by park base and other funding sources to promote integration and availability of datasets. Provide overall Network planning, training, and operational support for the awareness, coordination, integration of data and information management activities, including people, information needs, data, software, and hardware. Serve as Point of Contact for National Park Service database applications (NPSpecies, NatureBib, Dataset Catalog)	Aware of specific protocol and related subject(s) for all network projects. Aware of related protocols and projects. Aware of park resource information management and GIS Plans and capabilities. Aware of other related projects external to NPS.

Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
Network Data Manager (continued)	Ensure inventory and monitoring data are organized, useful, compliant, safe, and available	Coordinate internal and external data management activities. Assign and enforce data stewardship responsibilities. Review and approve all data acquisition plans, hardcopy and electronic field forms, and data dictionaries. Participate in development of Information Needs Assessments. Communicate with Crew Leader, Project Leader, I&M Network Coordinator, and Park GIS/Data Management office. Develop and maintain overall Network and individual Vital Sign data management operating guidelines and relationship to national standards and procedures. Develop and maintain the infrastructure for metadata creation, project documentation, and project data management. Create and maintain project databases in accordance with best practices and current program standards. Provide training in the theory and practice of data management tailored to the needs of project personnel. Develop ways to improve the accessibility and transparency of digital data. Establish and implement procedures to protect sensitive data according to project needs. Collaborate with GIS Specialists to integrate tabular data with geospatial data in a GIS system in a manner that meets project objectives. **Data managers will also work closely with the project leader to:** Define the scope of the project data and create a data structure that meets project needs. Become familiar with how the data are collected, handled, and used. Review quality control and quality assurance aspects of project protocols and standard procedure documentation. Identify elements that can be built into the database structure to facilitate quality control, such as required fields, range limits, pick lists and conditional validation rules. Create a user interface that streamlines the process of data entry, review, validation, and summarization that is consistent with the capabilities of the project staff.	

Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
Network Data Manager (continued)	Ensure inventory and monitoring data are organized, useful, compliant, safe, and available	Develop automated database procedures to improve the efficiency of the data summarization and reporting process. Make sure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of the project data. Ensure regular archival of project materials. Inform project staff of changes and advances in data management practices. Additional examples of the duties and responsibilities of the network data managers are listed in the I&M Program Vision and Organizational Framework document "Network Data Manager Overview of Responsibilities". **NOTE:** Data Managers with Prototype Monitoring Programs have the same basic duties and responsibilities as the network data managers but also are responsible for mentoring and training others and developing and testing new approaches to data analysis, synthesis, and reporting of monitoring results.	
Database Manager	Know and use databases and applications	Install, maintain, and support specific database software applications and NPS database applications. Work with Information Technology Specialists to resolve hardware and software issues.	Aware of NPS I&M Program generally. Aware of database tools and applications used by the I&M program. Aware of IT functions and capabilities at the park and network.
Curator	Oversee all aspects of the acquisition, documentation, preservation, and use of park collections	Know park natural resource collections Conduct accessioning, cataloging, legal, and other documentation of collections Manage collections databases Recognize objects needing conservation treatment Recommend and refer treatment to the appropriate facility Work with Network Data Manager to acquire and process data related to natural resource collections	Aware of NPS Museology Aware of database tools and applications used by the I&M program.

Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
Statistician or Biometrician	Analyze data and present information	Work with the Network Ecologist to analyze and report data according to established protocols. Work with the Network Data Manager to acquire and process raw data from databases and store derived data and information after analysis	Aware of capabilities, limitations, and applicability of statistical software applications Aware of database tools and applications used by the I&M program.
Network Ecologist	Integrate science in network activities	Ensure useful data are collected and managed by integrating natural resource science in network activities and products, including objective setting, sample design, data analysis, synthesis, and reporting. Assist with development and modification of monitoring protocols and inventory study plans. Work with the Network Data Manager to incorporate data management in monitoring protocols. Participate in the development of Information Needs Assessments based on the objectives of the project. Guide and/or perform statistical and other analyses of network data. Contribute to the synthesis and reporting of data and information. Provide guidance and support, to the extent possible, to extend Network standards and procedures to studies and data funded by park base and other funding sources to promote integration and availability of datasets.	Aware of park resource information management and GIS Plans and capabilities. Aware of database tools and applications used by the I&M program. Aware of Data Stewardship principles.
Network Coordinator	Coordinate all network activities	Ensure programmatic data and information management requirements are met as part of overall Network business. Communicate with Network staff, park staff at all levels, and other appropriate audiences to support and emphasize data management as a critical aspect of network business Work with Network Data Manager regarding data management policy and guidelines, budget, staffing, and training. Hold Network staff accountable for responsibilities involving data management.	Aware of database tools and applications used by the I&M program. Aware of Data Stewardship principles.

Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
I&M Data Manager (National Level)	Provide service-wide database availability and support	Provide services to receive, convert, store, and archive data in service-wide databases. Work with Network Data Manager to resolve local issues involving the access and use of inventory and monitoring databases. Provide training where possible. Design and maintain standardized, master databases for Servicewide planning, decision-making, and accountability (e.g., NPSpecies, NatureBib, Dataset Catalog, Database Template, GIS tools). Collaborate with networks to help develop overall data management vision and approach, and continual improvement of specific tools. Coordinate establishment of standards for naming conventions and content of data management plans and monitoring protocols. Promote collaboration and integration with other divisions and programs including the GIS community, fire program, air resources, water resources, geologic resources, etc. Facilitate coordination and collaboration among the parks and networks by providing examples of good database designs with flexibility to allow adjustments for different situations.	Aware of network contacts
Other End Users	Use and apply Network services and products	These 'information consumers' include park managers and superintendents, researchers, staff from other agencies, and the public. End users at all levels are generally responsible for providing necessary and requested feedback, review, and comments on various products in order to sustain and promote continuous improvement of network operations and services. End users are responsible for the appropriate use and application of data and derived products, e.g. by reading metadata. Administrative unit managers and program managers, in particular, are responsible for providing the money and scheduling the time required for project leaders and staff to meet stated data quality and other data management objectives.	Varies with each end user. Awareness is an outcome of end use. Awareness of the resources (time, expertise, and funding) required to develop and maintain an effective data and information management system that meets the business needs of the end user(s).

Appendix B: GRYN Folder and File Naming Conventions

Files are stored in a hierarchical folder structure that provides context to the information system. Folder structures are kept as flat as possible while providing for adequate classification of common project elements. The organization and naming of folders and files are concise and descriptive to promote understanding by users unfamiliar with specific projects and administrative functions.

Folder Structure

Computer drive letters G, I, and U are shared locations of hard drives or disk partitions on the GRYN file server. As corporate information resources these drives and their contents meet National Park Service Security standards for confidentiality, integrity, and availability. Controls exist for physical and electronic access to the file server and backup of all corporate and user data occurs daily, weekly, and quarterly. Quarterly backup media is stored off-site.

G:\ (data drive)

The data drive (G:) is GRYN's corporate library for GIS data, NPS Natural resource Databases, and NPS Service-wide Databases. Data sources are stored in various spatial data formats, including ESRI shapefile, coverage, geodatabase, and image file formats. The directory structure organizes the data primarily by spatial extent such as park-wide, network-wide, and statewide. Within a directory like "yell" (for Yellowstone) the data are organized in folders named by resource type or discipline, for example "climate", "fire", and "water".

The "Natural_Resource_databases" folder contains a variety of files in Microsoft Access format based on the National Park Service's Natural Resource Database Template (NRDT). GRYN staff members are investigating the long-term utility of the NRDT versus developing data sources strictly as ESRI geodatabases, or using a combination of geodatabase and NRDT.

The "Servicewide_databases" folder holds desktop versions of standard database applications developed for the entire Park Service. These include NPSpecies, NatureBib, Dataset Catalog, and NPSTORET. Some of these databases have counterpart online interfaces to NPS corporate master databases that reside on servers at the national I&M Program Office in Fort Collins, Colorado. Users must be aware of and account for the synchrony between local desktop or file server copies and master corporate databases. This is an issue when updating and using the database because users must ensure that changes to local data are communicated to the corporate database, and that the local copy of the data represents the complete content of the corporate database.

Top-level folders of the G:\ ('data') drive:

bica	national
Documentation	Natural_Resource_databases
grte	Servicewide_databases
gryn	Toolsets
ID	WY
MT	yell

Replicas of GRYN corporate datasets on local computers are stored in explicit locations on the local hard drive that indicate a copy of a master information resource. For example, GRYN corporate data stored on the workgroup file server's G: drive is stored locally at C:\NPS_Corporate_Data\G_Drive_Replica\

U:\ (User drive)

The user drive (U:) holds I&M staff workspaces and corporate files representing the business of the I&M Network. Each GRYN staff member with login access to the local computer network (IMWORKGROUP) is assigned a top-level folder named with their user name. User workspaces support daily operations to assemble and manipulate files and documents, prepare drafts and sections of documents, etc. The U: drive also contains folders and files specific to the business of the program, such as "Admin", "Inventory", and "Monitoring".

Top-level folders of the U:\ drive:

Admin	Monitoring
Data_Info_Management	Program_Web_Portal
General_Library	Users
Inventory	

Managing data and information resources across computer networks

Some GRYN personnel are not directly connected to the GRYN computer workgroup which is supported by a US Geological Survey Local Area Network. This prevents access to GRYN's shared disk drives. This is due to staff using non-NPS computer hardware or because these staff are located at offices served by a different Local Area Network. Staff working off-network and those without access to network drives must be aware of GRYN corporate folder structure and should follow the GRYN guidelines for file naming and storage. Using the following GRYN directory structure and naming convention on desktop and laptop computers facilitates information exchange.

The convention for local disks includes the following:

C:\NPS_Corporate_Data\G_Drive_Replica\

C:\NPS_Corporate_Data\I_Drive_Replica\

C:\NPS_Corporate_Data\U_Drive_Replica\

C:\NPS_Working_Data\

C:\NPS_Install_Files\

Local users should build the basic structure listed above and need to replicate and synchronize only those folders and subfolders relevant to the work they contribute for the GRYN Program. For example, an off-network employee working on the arid seeps and

springs Vital Sign will work in the following local directory: C:\NPS_Corporate_Data\U_Drive_Replica\Monitoring\Vital_Signs\Arid_Seep_Spring\ Subfolders under Arid_Seep_Spring should follow the corporate Vital Sign subfolders listed below. It is not necessary for this user to build the other vital sign folders and subfolders on the local computer. If users choose not to replicate the corporate structure on the local machine, they must communicate their local file management design and explicitly define how it relates to the GRYN corporate structure. This can be submitted in writing or verbally communicated to the GRYN data manager.

Information transfer can occur by several methods, depending on the circumstances. All off-network personnel can use e-mail to exchange files. Small files (less than 500KB) can be directly attached to an email message. Larger files or collections of files should be compressed to a target size of 3MB or less. Larger compressed files should be transferred using other methods, including physical transfer of files on USB storage devices and optical disk media (DVD or CD), or file transfer protocol (ftp). Those personnel using NPS laptop computer hardware can also transfer large volumes of information by physically connecting the laptop to a site on the USGS Local Area Network that serves GRYN. Those not using NPS hardware are prohibited by Department Of Interior security policy from directly accessing the USGS LAN.

The data manager assists users to reconcile and synchronize the contents of the corporate files on the file server and the replicas on desktop and laptop computers.

To gain access to GRYN information resources, off-network users request CD or DVD media containing copies of GRYN files, or connect NPS-property computers to the network temporarily to transfer files.

Synchronization of files between staff computers and the GRYN file server should occur frequently in order to maintain a current corporate library that is available to all GRYN staff, other NPS employees and for including in backup and security procedures. Synchronization is strongly recommended on a monthly basis or more frequently.

Vital Signs Information Management

The U:\Monitoring directory and its subfolders represent the strategic planning at GRYN to develop a consistent, simple, and comprehensive file management model that directly supports development of products like the Network Monitoring Plan and individual Vital Sign Monitoring Plan and Protocol documents, as well as web-interfaces targeted to a variety of audiences.

The structured design of folders, file naming convention, and file content format allow for functional development and maintenance of information resources as manageable components. The design provides for easy assembly of information products, both hardcopy and web-based. These products may be compiled manually at first, but will later be automatically generated based on content tags and markup language.

Information sources are named and filed according to this convention. Filing information sources in shared locations and using descriptive file name are important steps to develop a useful corporate information system. In cases where users are acquiring several files from another source, it is appropriate to file them in the best location and follow up as soon as possible with file naming that follows the GRYN convention.

Versioning and Change Procedures

The structure of the Vital Signs directories and files supports a version tracking system based on file name that avoids creating multiple subfolders. The master file in each Vital Sign subfolder is indicated by the word 'current' in the file name. When a master file meets the criteria for a version change (See below) it is copied to the same folder and given a different file name that replaces the word 'current' with the next available sequential version number, preceded by the letter 'v'.

For example, the Whitebark pine issues and threats document "isth_Whitebark_Pine_current.doc" is the current, master file that is supported by two previous versions. The most recent version is a file named "isth_Whitebark_Pine_v3.doc"

In addition to the file naming convention for version tracking, a reader can establish the timing of related documents by viewing the file properties, that include the date and time the file was created, modified, and accessed.

What constitutes a version change and how does GRYN track changes in documents?

The scope of this guideline is for documents produced in Microsoft Word, the NPS-standard documenting application. To address compatibility issues and user requirements on various computing platforms, documents may be published and distributed as Adobe Portable Document Format (pdf).

GRYN materials are developed at two levels – one level is draft, internal working documents leading to publication of an initial or subsequent products for distribution. The other level is a previously published, external document that compels changes to network operations. Examples of substantial changes requiring a new version include a change in monitoring objectives, sampling design, data collection, analysis, or distribution methods, or any aspect of the inventory and monitoring program that requires a process adjustment. This might be the result of information learned during a pilot implementation of a monitoring protocol. Document edits that do not require a separate version include text manipulation, grammatical and formatting changes, addition of supporting content not resulting in a process adjustment, etc.

It is necessary to track changes in draft, pre-publication documents, and track versions in published, updated documents.

GRYN content is developed with multiple levels of interaction and review. For example, Individual GRYN staff author documents which may be exchanged with others for review and edits. In order to establish and track the master copy of a working document, the

original author will usually act as Document Steward and be responsible for storing the master copy, keeping a current copy of the master document in the corporate GRYN file system, distributing the document for edits and review, providing instruction for reviewers/editors, receiving tracked changes(edits), comparing and merging edit copies into the master, managing version of the document, etc.

Example: The network coordinator emails a note to several people requesting review and edits to document "standard_MOU_v1.doc". The email includes the local network file path and file name as well as the attached document. The document has the track changes feature turned on. The email message includes instructions to use the track changes feature, how to provide edits and return the edited document. The procedures that follow depend on whether or not the collaborating individual has access to the local network.

Off-Network Collaborators: Each person saves the attached document to a personal working directory and includes their username and the word 'edits' in the file name. For example, the attached document "standard_MOU_v1.doc" is saved locally as "standard_MOU_v1_rdaley_edits.doc" Edits are made using the track changes feature, which allows Cathie as the Document Steward to compare and merge the changes from multiple review copies when they are returned.

ISSUE: Occasionally more than one person will need to review and edit a draft document at the same time.
SOLUTION: It is best to maintain a single edit copy of the network based document. If it is necessary for more than one person to review and edit the document at the same time, then when the second or subsequent users sees the message about "Read only – someone else is using this file", that user will make a copy of the file following the naming convention outlined above, i.e. "standard_MOU_v1.doc" is saved locally as "standard_MOU_v1_rdaley_edits.doc"

Document Identification Standards: Information in the footer of the document to help readers use the documents. Include the following:
Document Steward – name of a single individual responsible for the master document
File path and name
Date file originated
Date file last modified
Page # of ##

U:\Monitoring\Vital_Signs_ (Vital Signs directory structure - DRAFT)

Vital signs folders under the 'Monitoring' folder:

(This list represents the network's overall vital signs list, for some of which GRYN develops monitoring protocols.)

Algae
Amphibians
Aquatic_Invertebrate_Assemblages
Arid_Seep_Spring
Atmospheric_Deposition
Backcountry_Day_Use
Backcountry_Overnight_Use
Beaver
Biogeochemical_Flux
Birds_Of_Concern
Climate
Community_Alpine
Community_Aspen
Community_Cushion_Plant
Community_Riparian_Riverine
Community_Shrub_Steppe
Ecoli
<example_structure> (standard subfolders)
Exotic_Aquatic_Assemblages
Fire
Forest_Insect_Disease
Geothermal_Features
Geothermal_Water_Chemistry
Glaciers
Ground_Water_Quantity
Insects
Invasive_Plants
Land_Birds
Land_Cover
Land_Use
Large_Carnivores
Mesocarnivores
Native_Aquatic_Assemblages
Oversnow_Emmissions
Reservoir_Lake_Elevation
Seismic_Activity
Soil_Structure_Stability
Soundscapes
Stream_Sediment_Transport
Streamflow
Ungulates
Vertebrate_Disease
Visibility
Visitor_Use
Water_Chemistry
Water_Quality_Regulatory
Water_Temperature
Whitebark_Pine (example below)

Common (required) subfolder structure under each vital sign folder:

<vital_sign>
- !Overview
- Conceptual_Models
- Current_Historic_Monitoring
- Data_Management
- Issues_Threats
- Meetings
- Monitoring_objectives
- Projects
  - <specific project folder> *(replace with project name)*
    - !Bin
    - Agreements
    - Project_Deliverables
    - Proposals
- Protocol
- Resources
  - !Bin
  - Guidance_Documents
  - Images
    - Restricted
    - Unrestricted
  - Journal_Articles
  - Maps
  - Other
  - Popular_Articles
  - Presentations
  - Related_Protocols
  - Reports
  - Web
- Sampling_Design

Below is an example with explanation of common subfolder structure shared by all folders with a vital sign name. The example uses the Whitebark Pine vital sign name:

Example of common subfolder structure shared by all folders with a vital sign name:

U:\Monitoring\Vital_Signs\Whitebark_Pine\

!Overview

This folder contains the following documents:

Protocol Development Summary(PDS) *for all protocols scheduled for implementation by year 2009. The PDS briefly explains why this vital sign is important, lists specific objectives, outlines the basic approach to develop the protocol (e.g. borrowing from existing protocols), and information about who will develop the protocol, how long it will take, what it will cost, etc. Summarizing and formatting this information to share among the networks will enhance opportunities for collaboration and consistency. PDS's are for internal use to promote communication and collaboration.*

Example: Whitebark_Pine_PDS_current.doc

Vital Sign Summary *for those Vital Signs not scheduled for implementation by year 2009. The Vital Sign Summary describes the vital sign, provides background, and explains the justification for developing the vital sign plan and protocol.*

The "Technical Notes" from Phase 2 planning *are a source of information that relate to Parts 1 and 2 of the monitoring protocol document.*

Project Status Summary *lists by date those efforts and resources related to the Vital Sign, and may contain hyperlinks to materials stored in the 'Projects' subfolders for this Vital Sign.*

Workshops Summary *lists by date those workshops related to the vital sign.*

Change Log *lists substantive changes related to the vital sign.*

Conceptual_Models

Current_Historic_Monitoring

This folder contains a document that describes the who, what, when, where, and why for monitoring projects taking place in the network, but not directly by GRYN.

Data_Management

Issues_Threats

Meetings

Monitoring_objectives

Protocol

This folder contains a structured document compiled from master component documents stored in the other folders for this Vital Sign, in addition to Standard Operating Procedures stored in folder U:\Monitoring\SOPs. Versions of this document are compiled as needed for hardcopy publication and distribution.

Projects

<specific project folder> *(replace with project name)*

!Bin

Agreements

Proposals

Project_Deliverables

Resources

!Bin

Guidance_Documents
This folder includes documents concerning legislation, directives, policies, etc. related to this vital sign. This may include a html file with URL links to current information sources.

Images

Restricted
Contains images with known copyrights

Unrestricted
Contains images with no known copyrights

Journal_Articles
Contains peer-reviewed articles from scientific journals, generally in Adobe portable document format (PDF)>

other

Popular_Articles
Contains articles that are not peer-reviewed

Presentations

Related_Protocols

Reports

Web
includes cited copies of web content and a html file containing links to web sites related to this vital sign

Sampling_Design

<other subfolder(s) > *(add additional folders for this vital sign when the time is right and after discussion with GRYN staff)*

Best Practices for Folder and File Naming:

Use official/original file sources
Where possible, avoid storing local copies of files that can be accessed via web link to the entity or organization that keeps the master file. This prevents storing and using an outdated copy and relieves network staff from monitoring the source for updates. An example is a web site with federal or NPS forms that can be accessed when needed.

It is necessary to use GRYN corporate folders to store files that support a particular version of a GRYN product or document. For example, files containing current NPS policy and I&M guidance that support GRYN monitoring protocols must be archived with the protocol documents to avoid confusion as policies and protocols change over time. Likewise, data records and data sets and the subsequent use, analysis, and reporting of the data link to specific protocols. Since protocols will change over time, each preceding protocol is archived with supporting documentation, and all data associated with each version of a protocol is documented to support the utility and longevity of the data and resulting information.

Avoid spaces in file names

Avoid using special characters, other than underscore, in file names

Avoid uncommon acronyms, abbreviations and codes
Avoid codes that require any user unfamiliar with GRYN business to refer to another table or document for the code description.
Use recognizable abbreviations or spell it out.

Exceptions: Park Service Unit acronyms like GRYN, YELL, GRTE, BICA
Acronyms in common use within and outside the National Park Service and related to natural resource management.

For additional details and reference, refer to:
http://science.nature.nps.gov/im/apps/template/IM_DB_Naming_Recs_v3.pdf":
This file stored at this location on the GRYN server:
U:\Data_Management\Library\Guidance_Documents\ IM_DB_Naming_Recs_v3.pdf

Definition and Examples of file naming convention

The GRYN file name conventions are as follows:

For articles published in peer reviewed journals:
GRYN staff will comply with the terms set forth in each journal publication regarding use and distribution of journal articles.

Files are stored in the following folder:
U:\Monitoring\Vital_Signs\<vital_sign>\Resources\Journal_Papers\

The file name is composed of the following elements separated with the underscore character:

- last name of the first listed author
  If a publications has two authors:
  - after last name of first author include last name of the second listed author

  If a publication has more than two authors:
  - after last name of first author include "et_al"
- year published (if necessary use <Year>a etc. for multiple articles by the same author in the same year)
- journal title abbreviation. Once source for abbreviations: http://www.bioscience.org/atlases/jourabbr/list.htm
- article title {optional} - a reasonable portion of the title with limited use of arbitrary abbreviations
- file extension – a three-place file format extension.

syntax:
<Author>_<Year>_<Journal_title>_{Article_title}.<ext>
<Author1>_<Author2>_<Year>_<Journal_title>_{Article_title}.<ext>
<Author1>_et_al_<Year>_<Journal_title>_{Article_title}.<ext>

examples:
single author example: **Corn_2003_conserv_biol.pdf**
two-author example: **Haines_Pollock_1998_Environ_Ecolo_Stats.pdf**
more than two-author example: **Vos_et_al_2000_Env_Mon_Ass.pdf**

For reports such as status reports of inventory or monitoring work required as a deliverable from an agreement with a cooperator or a contract.
syntax:
<vital_sign>_<Author>_<Report_title>_<Year>

References not fitting into other categories

Directives folder:

Meetings folder:
<vital_sign>_<Date>_<Meeting_location>_<Key_group(s) or Key_Individual(s)>

Actions folder:

Decisions folder

Issues_Threats folder:

The following file name prefixes were considered for use by GRYN staff using as a convention to prefix document and file names. Since the directory structure provides context to such material, use of these prefixes is optional and generally will not be used.

File prefix	type of document or file
lit_	peer reviewed journal article
rep_	report title, e.g. a deliverable from a cooperator
ref_	reference documents
pres_	presentations such as PowerPoint
reg_	regulation
pol_	policy
exec_	executive order direction
mtg	meeting notes
act_	management actions
dcsn_	management Decision
isth_	issues and threats

Appendix C: Freedom of Information Act (FOIA) and Sensitive Data

From
Northeast Coastal and Barrier Network Data Management Plan
Marc Albert and Sara Stevens
November 2004

This appendix summarizes the laws and policy related to protected information about Park resources and the Freedom of Information Act (FOIA). It also describes the procedures for classifying and managing protected information from Inventory and Monitoring Program projects, as well as the procedures for responding to FOIA requests. Much of the material contained in this section is copied or derived from NPS Director's Order #66: Freedom of Information Act and the Protection of Exempted Information (Drafts 12-4-03 and 4-12-04).

C.1 Summary

The FOIA specifies a process through which all United States Government entities must respond to requests for information by any member of the public. FOIA and National Park Service policy require that NPS staff routinely make available information that is of interest to the public, including data regarding park resources and management. Resource information collected through the Greater Yellowstone Network, whether by NPS staff or partners, is intended to be available not only to parks but also to the public, and the routine dissemination of resource information is an important component of the Network Inventory and Monitoring Program.

However, information that could result in harm to resources may be withheld from public release. Four resource confidentiality laws and an Executive Order direct the NPS to protect information regarding the nature and location of certain sensitive park resources. One of these laws, the National Parks Omnibus Management Act, states that information that could result in harm to specific natural resources, including endangered or threatened species, may not be released to the public, and that records containing such information are exempted from release through FOIA.

Only a small subset of the information collected through the Inventory and Monitoring Program is likely to be considered protected. Nevertheless, all data sets and associated information from Inventory and Monitoring Program activities, including spatial data such as GIS files, should be assessed to determine sensitivity, and any protected information should be carefully managed to prevent its release. When publishing or posting resource information, or when responding to a FOIA request in collaboration with the regional FOIA officer, Inventory and Monitoring Program staff should try to ensure that only the protected information is withheld and not associated non-sensitive information.

C.2 Definitions Relating to Management of Protected Information

Endangered or threatened National Park System resources. For natural resources, this indicates a species or population that has been formally designated as endangered, threatened, a species of concern, or proposed for such a designation by the U.S. Fish and Wildlife Service or the National Marine Fisheries Service, or a similar designation by an appropriate state agency. A parallel definition exists for cultural resources.

Partners are individuals or entities that enter into cooperative or collaborative relationships with NPS for the purpose of achieving overlapping goals, where at least some goal is held in common by both the partner and the NPS. This relationship is documented through a letter, VIP appointment, general agreement, permit, contract, or some similar written arrangement. Partners can include both private entities and other federal agencies.

Protected information indicates information about a sensitive park resource that must not be released to the public according to any of the four resource confidentiality laws and the Executive Order applicable to NPS.

Principal Investigator in this context means the person primarily responsible for the implementation of an inventory or monitoring project, whether that person is a partner, is affiliated with a partner institution, or is an NPS employee.

Record includes all books, papers, maps, photographs, machine-readable materials, or other documentary materials, regardless of physical form or characteristics. Records are made or received by an agency of the United States Government under federal law or in connection with the transaction of public business and preserved by that agency as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the Government, or because of the informational value of the data in them (44 U.S.C. 3301).

Resource confidentiality laws refer specifically National Parks Omnibus Management Act (16 U.S.C. 5937), National Historic Preservation Act (16 U.S.C. 470w-3), Federal Cave Resources Protection Act (16 U.S.C. 4304) and Archaeological Resources Protection Act (16 U.S.C. 470hh).

Sensitive park resource indicates a park resource covered by the resource confidentiality laws that is considered susceptible to significant harm, theft or destruction, and about which information should be protected from public release.

C.3 Legal and Policy Framework for NPS Natural Resources Information Dissemination

C.3.1 Freedom of Information Act, 5 U.S.C. § 552

The Freedom of Information Act of 1966 (FOIA), amended in 1F6 to provide guidance for electronic information distribution (the amendments are commonly referred to as EFOIA) applies to records that are owned or controlled by a federal agency, regardless of whether or not the federal government created the records. FOIA is a broad disclosure law intended to establish a right for any person to access federal agency records that are not protected from disclosure by exemptions. Under the terms of FOIA, agencies make some non-protected records generally available for inspection and copying in public reading rooms and via the Internet. Other records are provided in response to specific requests through a specified process. The Department of the Interior's revised FOIA regulations (43 CFR Part 2, Subparts A through E [see 67 FR 64527]) and the Department's Freedom of Information Act Handbook (383 DM 15) can be accessed at http://www.doi.gov/foia/.

C.3.2 National Parks Omnibus Management Act, Section 207, 16 U.S.C. § 5937

The National Parks Omnibus Management Act (NPOMA) prohibits the release, under FOIA, of information regarding the nature and specific location of certain cultural and natural resources in the National Park System. Information prohibited from release includes the location of endangered or threatened species–specifically maps or narrative descriptions indicating site specific locations. The law also identifies conditions under which the Secretary may release this information.

C.3.3 National Park Service Management Policies

The NPS Management Policies (2001) explain the dual goals of the National Park Service with regard to information on resources–to withhold information that will put particular resources at risk and to expeditiously release information that does not.

C.3.4 Director's Order #66 Freedom of Information Act and the Protection of Exempted Information (Drafts 12-04-03 and 4-12-04)

The final Order will function as a supplement to the Department of the Interior FOIA regulations. It is intended to clarify internal NPS operational questions and responsibilities regarding procedures, signature authority, security requirements, and the relationship of paper and electronic records to FOIA and EFOIA. Also, the final Order will specifically address records concerning the location and nature of specific types of park resources that are prohibited from disclosure by the resource confidentiality laws. The draft of the Order states "In general, any federal agency that holds information about the nature and specific location of park resources that qualifies as protected information under the provisions of NPOMA must withhold that information from the public unless the Director of the National Park Service or designee determines that its release would:

1) further the purposes of the unit of the National Park System in which the resource is located
2) not create an unreasonable risk of harm, theft, or destruction of the resource

3) be consistent with other applicable laws protecting the resource–the expected Order will be accompanied by Reference Manual 66 which will give more detail (refer also to NPOMA)

C.4 Public Access to Network Inventory and Monitoring Data

According to NPS Management Policies and Inventory and Monitoring Program goals, each Network will make information on park resources readily available. In addition, the 1F6 amendments to FOIA require that all information that is regularly requested, except exempted records, must be made available to the public via reading rooms and the internet. The five internet portals through which information from Greater Yellowstone Network projects will be made available to the public are listed in Table C.1.

Table C.1 Greater Yellowstone Network Internal Portals for Projects

Name	Description of Content	For More Information
GRYN Web Site	Reports and other information on all Network projects as well as Network parks, operations and staff	http://www1.nature.nps.gov/im/units/gryn/
NPSpecies	Information on species in the National Parks, including all records generated through the I&M Program	www.nature.nps.gov/im/apps/npspp/
NatureBib	Bibliographic references that refer to National Park System natural resources	http://www.nature.nps.gov/nrbib/index.htm
NR-GIS Metadata and Data Store	Documents, maps, and data sets containing resource information from all sources, and their associated metadata	http://science.nature.nps.gov/nrdata/docs/about.cfm
Biodiversity Data Store	Documents, GIS maps, and data sets that contribute to the knowledge of biodiversity in National Park units	http://science.nature.nps.gov/im/inventory/biology/

Both secure and public interfaces are maintained for each of the databases associated with these portals (the NatureBib interface is in development currently), and the public will have access to all information in these databases except those records marked as 'sensitive.'

C.5 Classifying and Managing Protected Information

The procedures for classifying protected information and managing information about sensitive park resources can be summarized as follows:

- Network staff (Coordinator, Data Manager, and/or other designated staff) will ensure that all known potentially sensitive park resources are identified.
- Network staff will ensure that investigators working on Network projects understand that (1) all data and associated information must be made available for review by Network staff prior to public release in any

format, and (2) that any information classified by the NPS as protected should not be released in any format except as specifically coordinated with the NPS (see section 9.2.5.2.2).

- Network staff will identify all known potentially sensitive park resources to the principal investigator for each project.
- All known references to potentially sensitive park resources that are generated from each project will be identified to the Network by the principal investigator for that project.
- For each project, the Network staff will provide a complete list of all references to potentially sensitive park resources in each park to the park superintendent for review.
- Each superintendent determines which information should be protected.
- The Network staff will ensure that all protected information is properly identified and marked before uploading into Network or National databases, and before archiving the databases.
- Network staff will ensure that all references to protected information are removed or obscured in any reports, publications, maps, or other public forum. Following the standard for FOIA requests, the Network will segregate the non-releasable information and where practical will not withhold associated releasable information.

C.5.1 Classifying Protected Information

The classification of protected natural resource information from Inventory and Monitoring Program activities will be done on a case-by-case, project-by-project basis. According to NPOMA, if the NPS determines that disclosure of information would be harmful, information may be withheld concerning the nature and specific location of:

- endangered, threatened, rare, or commercially valuable National Park System resources
- mineral or paleontologic objects
- objects of cultural patrimony

The Federal Cave Resources Protection Act (16 U.S.C. § 4304) similarly authorizes the withholding of information concerning the specific location of any significant caves.

The Network will work closely with the investigators for each project to ensure that potentially sensitive park resources are identified and that information about these resources is tracked throughout the project. Network staff will be responsible for identifying all potentially sensitive resources to the principal investigator(s) working on each project. The investigators, whether NPS staff or partners, should develop procedures to flag all potentially sensitive resources in any products that come from the project, including documents, maps, databases and metadata. All records and other references to the potentially sensitive resources should be specifically identified by the investigator when submitting any products. Partners should not release any information before consulting with NPS staff to ensure that the information is not classified as protected. See section C.5.2.2.

Network staff should compile information about potentially sensitive resources from each project and forward it in the context in which it would be made available to the public (report, map, database etc.) to each appropriate park superintendent (or his or her designee). Each superintendent will determine whether or not to protect the information. For inventory reports, monitoring project reports, or other stand-alone documents, this process will be most efficiently conducted as part of the final draft review for each document. For information contained in other formats that will not have a discrete review process, Network staff will be responsible for flagging any potentially sensitive information and forwarding a request to the appropriate Superintendent(s).

The following guidance for determining whether information should be protected is suggested in the draft Director's Order #66 (the final guidance may be contained in the Reference Manual 66):

1) Has harm, theft, or destruction occurred to a similar resource on federal, state, or private lands?
2) Has harm, theft, or destruction occurred to other types of resources of similar commercial value, cultural importance, rarity, or threatened or endangered status on federal, state, or private lands?
3) Is information about locations of the park resource in the park specific enough so that the park resource likely could be found at these locations at predictable times now or in the future?
4) Would information about the nature of the park resource that is otherwise not of concern permit finding the resource if the information were available in conjunction with other specific types or classes of information?

5) Even where relatively out-dated, is there information that would reveal locations or characteristics of the park resource such that the information could be used to find the park resource as it exists now or is likely to exist in the future?
6) Does NPS have the capacity to protect the park resource if the public knows its specific location?

In the Greater Yellowstone Network, most information that may qualify as protected will pertain to rare species of plants and animals, including federal and state-listed species. The information that may be protected could include the location, density or abundance, or presence/absence of the resources in question. Specific examples are maps, narrative descriptions, or monitoring plot locations indicating site specific locations of species.

Information that is already in the public domain can be released. For instance, the return of condors to the Grand Canyon has been well documented by the press. If parties request site-specific information about where the condors have been seen, this information can be released. However, specific nest site locations must not be released.

C.5.2 Managing Protected Information

C.5.2.1 General Procedures

Any information that a superintendent determines should be protected will be removed by Network staff, or by partners with Network staff guidance, before publication or the posting of documents or other media in which the information is contained. Following the standard for FOIA requests, the Network will segregate the non-releasable information and where practical will not withhold associated releasable information.

The method used to withhold protected information depends on the nature of the particular park resource and the medium in which the information is contained. It is the responsibility of Network staff, with guidance from park superintendent(s), to determine the appropriate measures to withhold protected information. In the Greater Yellowstone Network, protected information is likely to refer to the presence or absence and location of rare species.

It may be appropriate to generalize location data in order to make an area large enough so that the public will be provided some information without learning the specific location of the park resource. This principle can be applied to text descriptions of locations, to text or coded data located on field data sheets or in databases, to GIS files, or to printed maps. In the case of databases, all references to any resource regarding which information is protected should be deleted or otherwise concealed in any publicly accessible version. For example, when providing location information, cutting off the last digits in UTM coordinates will make the location general enough in some cases. This could apply to metadata files associated with GIS data as well. An option for GIS-based displays or printed maps would be to increase the pixel size to the point that finding the object of interest is not possible.

Four of the databases for natural resource related information from the Inventory and Monitoring Program–NatureBib for bibliographic references, NPSpecies for species records, and the Biodiversity Data Store and NR-GIS Data and Metadata Store for documents, GIS maps, and data sets–are equipped with the capacity to mark protected information when records are being uploaded. All records that are marked ‘sensitive’ upon uploading will only be available through the secure applications. Thus, access to information on sensitive park resources will be limited to NPS staff or partners who have signed a confidentiality agreement and procedures regarding the release of protected information can be provided along with access to the databases. It is critical that the Network implement quality control and quality assurance measures to ensure that anyone uploading records into these databases will know the procedures for identifying and entering protected information.

Precautions should be taken to avoid inadvertent releases of protected information. Examples of inadvertent releases are the use of protected information in the development of NPS interpretive and public information programs or the inclusion of protected information in National Environmental Policy Act documents.

C.5.2.2 Procedures for Working with Partners

Network staff must work with any partners that are collecting or reporting information from Inventory and Monitoring Program projects to ensure that:

- all of the records and other information associated with projects are submitted to the NPS
- protected information is identified as described in section C.5
- protected information is withheld from public release
- the NPS and the partner have a signed agreement including a confidentiality clause
- specific procedures for review of information that may be reported by partners are established (This consultation must occur before the partner releases potentially sensitive information to any outside party, whether as part of a publication, posted to a website, or pursuant to a FOIA request or any other request.)

C.5.2.2.1 Partner Agreements

The rights and responsibilities of the NPS and partners regarding potentially sensitive park resources should be stated clearly in any Cooperative Agreement, contract, Interagency Agreement, VIP agreement, or other written confirmation of a working relationship. Each agreement should address the following:

- Clarification of the ownership of data and associated information–the following text, adapted from guidelines developed by Acadia National Park, is recommended: "All associated data (including, but not limited to field notes, maps, slides, photographs, charts/graphs, tabular and GIS data with associated metadata) are required to be submitted to the Network annually and are owned by the National Park Service."
- All known potentially sensitive park resources should be named. Since the agreement itself is a public document only the names of the resources should be provided, not specific information about their distribution or abundance, in case that information should be protected.
- The procedures for the classification of protected information should be summarized, especially with regard to cooperation between the Network staff and the project staff (as described in section C.5.1).
- A requirement for the partner to withhold protected information and to consult with the NPS before releasing any information on sensitive park resources should be stated. Thus, the document should include a confidentiality agreement.

If agreements do not contain specific responsibilities for the NPS and the partner regarding protected information, then Network staff must work with partners to institute the appropriate procedures. Note that Federal ownership of information means that the information is subject to public release through FOIA.

C.6 Responding to FOIA Requests

When a Network or a Network park receives a specific FOIA request for Inventory and Monitoring Program information, they will handle it according to standard Department of the Interior and NPS procedures, following the Department's Regulations, the Department's FOIA Handbook, and the NPS Directors Order #66 and Reference Manual 66. General information and web links regarding FOIA are located in section C.3.1. (The NPS Northeast Region's FOIA Program Officer is currently Annette Sasso, and the Regional FOIA Officer is currently Edie Shean-Hammond.)

The procedures for responding to FOIA requests for Inventory and Monitoring Program-related information can be summarized as follows:

- When a request is received by a park or the Network, it is the responsibility of the NPS official recipient (the superintendent for a park, or the Network Coordinator) to comply with FOIA. Notification of the request is forwarded to the regional FOIA program coordinator and the request is logged into the Electronic FOIA Tracking System.
- The request is 'perfected' (finalized) by estimating fees and determining the requestors payment limit or request for payment exemption; from this date the NPS has 20 days to respond.

- The requested records are compiled by the recipient office, noting any records that contain protected information and thus are exempted from release. The recipient must consult with the NPS solicitor regarding any exempted records.
- When a record contains both exempt and nonexempt material, a reasonable attempt should be made to segregate and release nonexempt information.
- A draft response to the request containing the requested records along with an explanation of any records that have been withheld should be submitted to the Regional FOIA Program Coordinator, then forwarded to the Regional FOIA Officer and finally to the Regional Director for signature and release.

C.6.1 Inquiries for Information and/or Records

Upon receipt of a request, the FOIA Officer will make a determination as to whether it is subject to FOIA (i.e., a request for records) or merely a request for information. If the requester seeks an answer to a specific question, or an explanation of policy, procedures, or a Departmental action, DOI is not required to process the request under FOIA. Nonetheless, the FOIA Officer should refer the request for information to the appropriate office for response in a timely manner.

C.6.2 Exemptions from Release of Records through FOIA

Nine exemptions and three special law enforcement record exclusions permit the withholding of sensitive or confidential information from release through FOIA. Although the NPS does not rely on any particular exemption, the one most likely to be used in the context of park natural resources requires withholding records that are prohibited from disclosure by another statute. Four resource confidentiality laws and one Executive Order direct the NPS to protect information regarding the nature and location of certain sensitive park resources. These include the Federal Cave Resources Protection Act and the National Parks Omnibus Management Act (NPOMA), which requires that information potentially harmful to particular natural resources, including listed endangered or threatened species, be withheld from public release.

In some instances, acknowledgement that a particular resource exists at all in a park may reveal too much information. In such cases, a response that neither confirms nor denies the existence of such records may be appropriate in reply to a FOIA request. Such a reply is known as a Glomar response.

C.6.3 Release of Records through FOIA for Projects with Partners

FOIA dictates that once an agency has shared records with any party outside the federal government without a pre-release agreement, it must make the records available to any and all other parties who request them. This provision is referred to as the “release to one, release to all” rule.

When published research findings are produced under a grant or other Federal assistance, including funding from the Inventory and Monitoring Program, and the findings are used by a bureau in developing an agency action that has the force and effect of law (e.g., a policy or regulation), the research data related to such findings are considered agency records even if they are in the possession of the recipient.

C.6.3.1 Procedures for Working with Federal Agency Partners

In general, any federal agency that holds information about the nature and specific location of park resources that qualifies as protected information under the provisions of NPOMA must withhold that information from the public unless the Director of the National Park Service or designee determines that its release would:

1) further the purposes of the unit of the National Park System
2) not create an unreasonable risk of harm, theft, or destruction of the resource
3) be consistent with other applicable laws protecting the resource

When another federal agency informs NPS that it has received a FOIA request regarding information that the other agency holds about park resources, NPS first assists the agency in determining whether the requested records fit within the definition of protected information. The agency must withhold the information pending action from the NPS. NPS will ask the agency to forward the FOIA request to the NPS with either: 1) a preliminary recommendation that the information be withheld; 2) a preliminary recommendation that it be released; or 3) a statement that the agency will not be making a recommendation whether the information should be released.

The NPS will make its determination about what information, if any, is to be withheld based on information the NPS receives from the agency, the requester, any other party that it consults, and its own inquiry into whether the information can be released under the provisions of NPOMA.

C.6.3.2 Requests for NPS Records Held by a State Agency or Partner

Before sharing information with state employees, whether from state agencies or state funded universities, NPS must be aware that those state employees may be obligated to release information in their possession to any party requesting it because state freedom of information or sunshine laws require such release. In states with Freedom of Information laws that allow the withholding of certain types of information, it may be possible that state employees would have the authority to enter into contractual agreements with NPS to withhold protected information. NPS must not share protected information with any state employee where state laws require the release of all information in state records.

C.6.3.3 Requests for Information Received by NPS from Non-NPS Entities

The NPS cannot guarantee confidentiality of information received from any non-NPS entity. Once NPS receives information from others, its treatment of the information is governed by FOIA. Such information must be released in response to a FOIA request if it does not qualify as protected information. The NPS must, however, withhold any information it receives that does qualify as protected.

Appendix D: Using the Global Positioning System and Portable Data Recorders

Contents

The Global Positioning System (GPS) and the associated data collection equipment and data processing software are important tools for natural resource data acquisition and management. Poorly organized GPS efforts waste time and money, while thorough and well-planned projects with well trained GPS operators result in efficiently acquired, high quality data that forms the basis of credible inventory and monitoring projects.

> It is the responsibility of each person conducting business for the Greater Yellowstone Inventory and Monitoring Network to consult and follow the most current procedures and guidelines in three information sources:
>
> 1. National Park Service standard GPS operating procedures: *Field Data Collection with Global Positioning Systems Standard Operating Procedures and Guidelines, 02/10/2004*.
> 2. This GRYN GPS/PDR Operating Guide
> 3. Individual Vital Sign Standard Operating Procedures for Data Management or individual inventory study plans.

This Greater Yellowstone Network Standard Procedure Guide for Data Collection Using the Global Positioning System and Portable Data Recorders supplements the National Park Service's *Field Data Collection with Global Positioning Systems Standard Operating Procedures and Guidelines, 02/10/2004*. Readers should be familiar with and follow both sets of guidelines as well as individual inventory project criteria and Vital Sign monitoring protocols that may contain additional detailed procedures. This document outlines Best Practices to effectively collect feature location and feature attribute data for the Greater Yellowstone Network Inventory and Monitoring Program. It is not an instruction manual on GPS or PDR theory, equipment, or software, but includes some details about steps that should be performed on every project involving GPS and field data recording. Several references are listed that provide more specific information on all aspects of the GPS.

It is important for everyone involved with National Park Service inventory and monitoring projects to understand the capabilities and limitations of GPS technology in order to use it effectively. The decision to include GPS with other resource management information technology depends on how specific monitoring protocol or inventory criteria relate to existing data and information management tools. For example, it is not appropriate to spend hours or days in the field collecting GPS data when the same features can be adequately mapped in the office using a Digital Orthophoto Quadrangle as a background image in a GIS application.

Effective use of GPS for resource information management involves shared responsibilities that include strong communication and understanding by each person involved in the collection, management, and use of data resources. In order to promote understanding and inform good decisions about using tools like GPS, field crew members and crew leaders should receive briefings from higher level supervisors, project managers, and data managers at least once per season about how their efforts fit with park and network management objectives, National Park Service and Department of Interior Policies, and federal government requirements, including those of congressional offices, the Federal Geographic Data Committee(FGDC), etc. Inventory and Monitoring project managers should have a good understanding of resource information management issues and requirements and be aware of the challenges and limitations of field data collection, including the use of GPS. This is achieved by detailed and regular briefings and/or accompanying field crews to perform data collection at least once each season. Table 1 provides a framework for the roles and responsibilities involved in using GPS for inventory and monitoring projects.

Table 1. Data Stewardship Framework for GPS and Field Data Collection Activities

Position or Office	Primary Role(s)	Primary Responsibilities related to GPS & Field Data Collection	Recommended Awareness Level
Field Crew Member	Collect and Record Data	▪ Obtain training in GPS theory and equipment. ▪ Read and follow all three levels of NPS GPS Operating Guidelines. ▪ Communicate with Crew Leader, Project Manager, and Data Manager. ▪ Perform regularly scheduled data transfer and backup data for further processing. ▪ Review, validate, and correct field data. ▪ Assist with data documentation.	▪ Good understanding of specific protocol and related discipline. ▪ Some awareness of any related protocols. ▪ Briefed on Network Program, NPS I&M Program, and relationships with e-Gov and other agencies efforts and initiatives.
Field Crew Leader	Supervise crew	▪ Ensure crew members receive GPS training and briefings. ▪ Read and follow all three levels of GPS Operating Guidelines. ▪ Communicate with Crew Members, Project Manager, and Data Manager. ▪ Ensure data is regularly transferred, backed up, verified, and entered into the appropriate NPS database(s). ▪ Assist with data documentation.	

Position or Office	Primary Role(s)	Primary Responsibilities related to GPS & Field Data Collection	Recommended Awareness Level
Project Leader	Supervise or Advise Field Crew Leader. Oversee and coordinate the project.	▪ Ensure crew leader receives pertinent training and briefings. ▪ Read and follow all three levels of GPS Operating Guidelines. ▪ Communicate with Crew Leader, Data Manager, and I&M Network Coordinator.	
I&M Network Coordinator	Supervise or Advise Project Manager. Coordinate all projects in general.		
Network Data Management Office	▪ Ensure inventory and monitoring data is organized, useful, compliant, safe, and available. ▪ Coordinate with Network parks, other Networks, and regional and national I&M and Natural Resource Programs as well as other agencies and entities.	▪ Assign and enforce data stewardship roles and responsibilities. ▪ Review and approve all data acquisition plans, hardcopy and electronic field forms, and GPS/GIS data dictionaries. ▪ Communicate with Crew Leader, Project Leader, I&M Network Coordinator, and Park GIS/Data Management office. ▪ Develop and maintain overall Network and individual Vital Sign GPS Operating Guidelines and relationship to national standards and procedures.	
Park GIS/Data Management office	▪ Coordinate and integrate local GIS and resource information management with Network, regional, and National standards. ▪ Support park management objectives.		

Information Resources Lifecycle

Data collected using GPS-enabled equipment represents all or part of the acquisition stage of an information resources lifecycle that includes several other stages (see Table 5.1 in GRYN Data and Information Management Plan). Data collection using the same equipment may also relate to the data maintenance stage of the Lifecycle. Whether GPS equipment includes data logging functionality, or Portable Data Recorders (PDR's) are GPS-enabled, it is most efficient to use a single device to collect and store data about both the location and characteristics of features for inventory and monitoring projects. The process and methodology used for acquisition planning, data collecting, and post-processing incorporate several aspects of data management, including quality assurance, data storage and organization, and data stewardship. To promote data quality and simplify data management, the Greater Yellowstone Network expects to use electronic data logging equipment for some data acquisition. However, parallel or complementary use of hand written data sheets and field notes will remain important for data collection activities.

Mission Planning and Preparation

Information gathering and mission planning are essential requirements that help protect investments in field data collection efforts that result in data that meet the requirements of the inventory or monitoring project. It is a chance to anticipate and prepare for the challenges as well as take advantage of the opportunities related to field data collection.

Read the monitoring plan and protocol for the vital sign or the inventory project plan. It provides the context and details for the overall effort and approach, which may include using the GPS and GPS-enabled equipment.

Read National Park Service standard GPS operating procedures: *Field Data Collection with Global Positioning Systems Standard Operating Procedures and Guidelines*.

Obtain training on GPS theory and on the use of GPS equipment. This may include commercially or academically available training and GRYN-sponsored training.

Obtain training and experience in map reading, using a compass, and orienteering because battery powered electronic GPS equipment is not fail-proof and should not be relied on as the only means of navigation.

Work with the network data manager to identify and become familiar with the database(s) that will store and manage the results of the GPS data collection. In most cases the data will ultimately reside with related data in one or more corporate databases managed by the National Park Service, other federal agency databases, national park databases, and I&M Network databases.

Work with the Network data manager and project leader to design field forms and electronic data collection forms (e.g. Trimble© Data Dictionary, ArcPad© application, or other). The design of the form(s) is determined by the inventory and monitoring objectives, their protocols, and existing or co-developed database structures. For example, know and use existing database field names in forms rather than locally invented names. **All data dictionaries and field forms will include the Vital Sign Monitoring Plan or inventory study plan title, version number and publication date.** This information is a critical element of required metadata for all data

collected based on a given monitoring plan or inventory study plan. The NPS Natural Resource Database Template (NRDT) Data Dictionary documentation and database design structure is the appropriate starting point for database development. The NRDT sample databases and the I&M web-based monitoring protocol clearinghouse should also be consulted for existing databases that can be adapted to serve GRYN data management requirements.

Obtain as much information as possible about the field site(s) through site reconnaissance, dialogue with individuals who are familiar with the site and topographic or other maps.

Use GPS planning software such as Trimble Planning Software (www.trimble.com) or equivalent to predict satellite availability, account for site visibility obstructions, and determine the best observation periods related to GPS receiver settings, as well as predict hours which the field crew can work on site under favorable conditions for satellite reception. Print tables and graphs from the planning software to take to the field. Reliable results from planning depend on a current almanac file. Current almanac files can be downloaded from a recently used GPS unit to the planning software. If the GPS unit has not been used for more than three weeks, first acquire an updated almanac by powering on the unit outdoors. The almanac file can take up to 15 minutes to refresh.

Practice collecting and post-processing data in real-world conditions before undertaking official data collection missions. Expect to learn a lot during initial field sessions and post-processing efforts, and expect to learn some little tip or trick on every subsequent data collection mission.

Understand the power requirements of the equipment and plan for the batteries and other power sources (battery recharging, 12V DC (auto) adapters, etc) necessary to support the duration of your mission.

Document and communicate any and all deviations from these procedures along with any events or circumstances from the field that affect the viability of using these procedures. This feedback is necessary for data documentation and to continually improve these procedures.

Include in electronic field data forms (e.g. Trimble GPS data dictionaries) only those parameters or attributes specified in the monitoring protocol or inventory criteria that require on-site measurement, observation, or description. This reduces data collection cost and promotes data quality. Other required attributes for a given database will be incorporated with the data following field work. This approach allows field personnel to focus on protocol-specific data collection using uncomplicated forms that include pick lists of valid attribute values. It also saves space on portable electronic data recorders and helps to enforce proper data flow and prompt verification of field data. For example, don't include or require the entry of administrative data such as state, county, or quad names. These known geospatial attributes are easily joined later using a GIS and unnecessarily take up field time and PDR space.

Field Data Collection

Only collect data with equipment and field forms that are approved by the GRYN project leader and GRYN data manager.

Consider using an external antenna with a ground plane where appropriate to maximize satellite reception. External antennas can be mounted above a person's head and body to provide better reception and remove the sensor from internal noise of the GPS receiver.

Always use a restraint system (neck/wrist lanyard, hand strap, clamp, etc) to secure the equipment and prevent accidental damage or loss.

Verify or set GPS configurations. Do this at the beginning of each session unless you were the last person to use the equipment, in which case you probably already know the settings. Use proper receiver settings from NPS *Field Data Collection with Global Positioning Systems Standard Operating Procedures and Guidelines.* This includes PDOP, SNR, elevation masks, minimum satellites, and antenna height.

Distinguish between the navigation and data collection portions of the mission and use appropriate GPS configurations for each. For general navigation use more forgiving settings. For precise navigation in the vicinity of a target location and for feature data collection, use settings that yield the most precise location possible while still allowing the receiver to record positions. When field conditions, including terrain and canopy, and/or timing conditions, such as fall and winter missions, prevent reception using NPS recommended settings, use the best possible settings that permit the receiver to record positions. In all cases perform differential correction during post processing to improve the precision of the location data.

With Trimble GeoExplorer equipment, use only one rover file per day for each data dictionary. (In some cases a GPS operator may support integrated monitoring using more than one data dictionary, but preferably the integrated data elements will reside in a single data dictionary.) This minimizes post-processing time required to merge multiple field files.

For extended backcountry missions over multiple days, the investment in data collected is at risk from accidental loss and a field backup solution is desired but sometimes not practical depending on the PDR equipment. Whenever possible, back up field data at least once each day to a secondary device.

Post Processing

The Greater Yellowstone Network requires all GPS data to be post processed, including differential correction and data verification, preferably by the individual(s) who collected the data. It is important to perform the differential correction and verification as soon as possible following the field work (strongly preferred within two weeks) in order to maintain and apply the knowledge and memories of the field crew to the data.

What about Real Time Differential Correction? Due to predominantly mountainous terrain in and around the parks and the lack of coverage by land based beacons transmitting the differential correction signals, most missions will not benefit from real time differential correction. Exceptions include operating in areas covered by real time signals and subscriptions to satellite based real time signals.

Processing location and attribute data once the equipment and personnel return to the office after a data collection mission involves several steps that serve key data management requirements. Processing stages include organizing and securing the data, improving the accuracy of feature locations (differential correction), verifying and making corrections where necessary, and

exporting the data to GIS and/or other required formats. Once the data makes it through post processing it can be merged with related data using GIS and RDBMS, become part of larger datasets, and enter the realm of corporately managed data in relevant database structures and at administrative levels including park and national databases.

Establish a PC file structure
Use a standard file structure to manage the various stages of file processing. All files should be stored on a disk that is backed up daily. Subfolders for each data collection project should accommodate field (rover) files, base files, differentially corrected files, validated files, exported files for incorporating in a GIS, and any other required export formats.

Transfer PDR Data
Use Trimble's Data Transfer program to copy the rover files from the GPS receiver to the local PC. Be aware that rover files and differentially corrected files must be stored in the same folder so they can be opened together in Pathfinder Office's Map View for a visual representation of the differential correction results. Comparing uncorrected and corrected data is not mandatory but it is useful in building trust in and understanding the process.

Look at (don't touch) rover file(s)
While it doesn't make sense to edit rover files that aren't differentially corrected, you should look at and understand each stage of the data so you know what is happening during the process. Looking at the 'raw' rover data provides visual verification that the rover file contains the data you expect to see and is adequate for further processing. You probably also verified this in the map screen on the GPS receiver or portable data recorder, so it should look similar (nicer and bigger) on the PC screen.

Perform Differential Correction
Use Trimble's Differential Correction Utility to improve the precision of the features.

Use the closest base station provider if possible. If the closest station's data is unavailable or the base files will not transfer, try the next closest station. Call the GRYN data manager if you cannot obtain any base data for your rover files. Base station operators might archive the station data after two or three months, so it is advised to correct (and validate) your data immediately – within a few days of field work.

Verify and correct the differentially corrected data
Many times the differentially corrected data requires additional edits to fix missing or extraneous positions and correct attribute values. Differential Correction in the previous step uses available positions and geometry to improve the location accuracy of your feature data. The most important part of post processing is verification of the feature types and attributes by the person who collected the data. It is especially important to verify and "manually correct" for potentially missing positions during periods of poor satellite signal reception when the GPS receiver may not have logged enough positions to describe the feature alignment or location.

IMPORTANT: Prior to performing any edits to the corrected file make a copy in the folder that will hold validated data files. Depending on the feature type(s) and number of attributes, verification and edits can take a substantial amount of time. Save often during lengthy edit sessions. Remember, this is the last best chance for verifying and improving data quality for

data that is expected to support inventory and monitoring objectives over several years or decades.

One goal of verification is to delete errant positions that result in spikes or 'spaghetti' in line and area features (point features generally don't need to be edited due to their spatial simplicity). Spikes in line and area features are usually caused by unavoidable multi-path signals, and messy spaghetti of overlapping lines are caused by collecting too many positions in too short a distance along the feature – this can be avoided by using the Pause/Resume key on the GPS receiver while collecting positions. Another part of data verification is to check, and change if necessary, the attribute values for the features.
The features should represent a mapped model of what exists on the ground but should not be 'perfect'. That is, don't zoom in so much that you spend hours fine tuning one feature. Use the scale factor displayed in the upper left of the Pathfinder Office Map View window to relate to the purpose of the data. Data is expected to at least meet National Map Accuracy Standards for maps on publication scales larger than 1:20,000 scale (http://rockyweb.cr.usgs.gov/nmpstds/nmas.html). Sometimes you need to zoom in at larger scales (smaller areas) to find and fix the spaghetti and spikes in your features, but don't zoom to 1:100 scale and spend a lot of time there. Usually with position log intervals of 1 second or greater, you can see most of the 'messy stuff' at scales less than 1:2000. However, pay extra attention to data that represent places where you collected several positions in a line feature over a short distance, or stood in one place without pausing the collection of positions. These are circumstances where the data tends to get 'messy' and needs to be edited to clean it up. As you become familiar with the relationships between equipment operation, your movements in the field as the 'platform' for the GPS equipment, and the resulting data, you will collect and produce cleaner data that takes less post-processing effort.

Understanding map scale. It is helpful to know that generic use of the term 'scale' in many natural resource discussions contrasts with explicit numeric ratios (map scales) that relate distance between features on a map to the true ground distance represented. For example, a map with scale 1:100 (**0.01**) means that one distance unit on the map represents 100 of the same distance units on the ground. Compared to a 1:100,000 (**0.00001**) scale map, the 0.01 scale ratio is larger and the map depicts a smaller area. So, maps with small scale ratios represent larger areas than maps with large scale ratios and vice versa. This differs from the more intuitive understanding of scale in a statement like "The project involves a large scale analysis of several ecological parameters," which is normally interpreted as a project involving a large area.

Open the corrected file in Pathfinder Office. (Optionally and carefully, you can open both the rover file (.ssf) and the corrected file (.cor) for the same Trimble data to compare the two. If you do this, it is recommended that before you edit the corrected features, you close all open files and reopen only the corrected file to avoid confusion about which file is which because you want to be sure to edit the corrected file rather than the rover file.)

IMPORTANT: Prior to performing any edits to the corrected file, create a copy of the corrected file in the folder for validated files. This helps keep track of the processing stage of the data file, and effectively leaves a 'backup' of the differentially corrected file in case there is a problem during the edit session. The "validated" folder will hold all the data that you've spent significant time processing and cleaning up.

From the "Data" menu in Trimble Pathfinder Office, make sure "Feature properties" and "Position Properties" are checked so they are visible on screen. If you don't see them, use the "Window" menu to tile or cascade the windows and then arrange them agreeably. You can also make the 'timeline' visible if it helps. Keeping the feature properties window bigger than position properties lets you see all or more of the attribute values and distinguishes between the windows so you don't accidentally delete a feature when you meant to delete a position. Remember that you can undelete until you save the edits. If you really mess up the edits, you can always start over with a fresh copy of the corrected file or by differentially correcting the rover file again.

Use the tools for zooming and panning to navigate among and along the features to examine them for spikes and spaghetti. Remember not to zoom in too much in an attempt to make the features perfect. Try a zoom scale of 1:2000. You may need to zoom in more on a certain messy area, but remember to zoom back out again to validate other parts of the feature.

Use the arrow tool to select features and positions.

Look at the position properties and feature properties windows for information and clarification.

Once the corrected data is validated and edited make sure it is saved to the folder containing validated files for that project.

Copy the validated data to a network drive or other backup device before proceeding. Once the same file is on both the local PC and the backup device (always try to keep each file in two places), it is safe to delete the rover file from the PDR in preparation for the next collection session.

Inform the project manager and data manager about the validated status and file location.

Work with the project manager and data manager to document the data fields, value domains, and other required metadata elements according to FGDC Metadata Standards.

Creating the GIS Data

Exporting files from GPS format to GIS format is an important step that nearly completes the GPS data acquisition portion of the information resources lifecycle, and provides an important opportunity to add value and utility to the data by including feature level metadata with the exported files.

Use the Trimble Pathfinder Office export utility to output the appropriate GIS format based on discussion with the network data manager. The network requires the following GPS Generated Attributes with exported data:

Generated Attributes for all feature types:

PDOP, Correction Status, Receiver Type, Date Recorded, Update Status, Data File Name, Total Positions, Data Dictionary Name (As specified in the Mission Planning section of this document, the data dictionary name represents and links to the specific version of a Vital Sign monitoring protocol).

Generated Attributes for point features:
Height, Vertical Precision, Horizontal Precision, Standard Deviation

Generated Attributes for line features:
Length (2D), Length (3D), Average Vert. Precision, Average Horiz. Precision

Generated Attributes for area features:
Area (2D), Perimeter (2D), Perimeter (3D), Average Vert. Precisions, Average Horiz. Precision

Verify the GIS data generated from the GPS data source to make sure the conversion was successful and the spatial, attribute, and GPS metadata elements exist. If problems are discovered, reexamine the GPS data source and the procedures used to convert to GIS format. If necessary, consult with the network data manager or park GIS Specialist to solve the problem.

Cleaning Up Unneeded Files
If the GIS data from the PDR/GPS source is complete and backed up in at least one other location, archive the corrected and validated GPS data files and delete the rover files, base data files, and other files generated during the GPS portion of post-processing. Provide the archived GPS data to the network data manager in a standard compressed and zipped format, i.e. Winzip.

References

References are available online or by request from GRYN Data Manager.

Field Data Collection with Global Positioning Systems Standard Operating Procedures and Guidelines, NPS, 02/10/04
http://www.nps.gov/gis/fm/documents/Field_Data_Collection_with_GPS02-04.doc
GRYN Server: U:\Data_Info_Management\Acquisition\Field_Data_Collection_with_GPS02-04.doc(6/29/04 download)

DRAFT GPS Data Accuracy Standard, US Forest Service, March 14, 2000
http://www.fs.fed.us/database/gps/gps_standards/GPS_Data_Standard.pdf
GRYN Server: U:\Data_Info_Management\Acquisition\USFS_Draft_GPS_Data_Standard.pdf (6/29/04 download)

National Park Service GPS for GIS Workflow, www.nps.gov/gis/gps/gps4gis/

Natural Resource Database Template, Inventory and Monitoring Program, National Park Service, http://science.nature.nps.gov/im/apps/template/index.htm

Protocol Database, Inventory and Monitoring Program, National Park Service, http://science.nature.nps.gov/im/monitor/protocoldb.cfm

GPSToolBox, Alaska Support Office, National Park Service, http://www.nps.gov/gis/gps/aksogps/

GPS to GIS Procedural Handbook & Reference Guide *VERSION 6.0,* San Juan National Forest, US Forest Service, 12/08/2003.
GRYN Server: U:\Data_Info_Management\Acquisition\USFS_San_Juan_gps_gis_v6-0.pdf

Class Notes – Design, Capture, Edit, And Integration Of Resource-Grade Trimble® GPS Data, Coconino National Forest, US Forest Service, January 2004

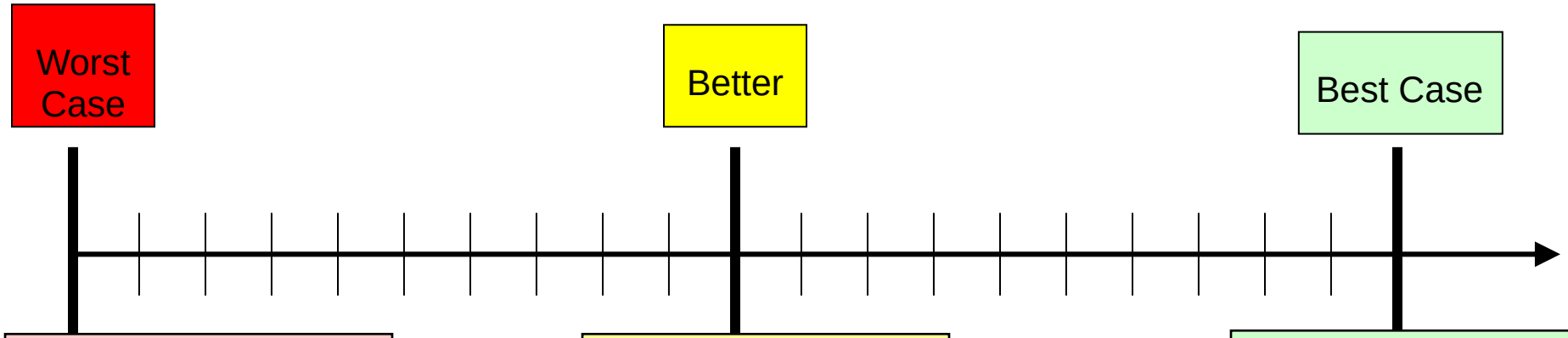
National Park Service Geography and Mapping Technologies, www.nps.gov/gis

National Park Service Global Positioning Systems, www.nps.gov/gis/gps/

USDA Forest Service GPS, www.fs.fed.us/database/gps/

Trimble Navigation Limited, 2004, www.trimble.com

The Continuum of GPS Data Collection



GPS user works alone.

No mission planning.

No data dictionary or dictionary does not reflect monitoring protocol.

User does not promptly download, correct, verify, edit, and export the data.

Data is not communicate/shared with others.

Data is not related to National Park Service protocols.

GPS user works with others occasionally.

Some mission planning is done, if convenient.

User downloads and looks at data sometimes.

Data is shared with others at the Unit.

Data is related to NPS protocols in some ways, not in others.

GPS user maintains a network of contacts.

Mission planning is part of the work flow.

User follows through with downloading, differentially correcting, editing if necessary, and exporting.

Data is collected, processed, and shared to meet or exceed agency standards.

The user has a good understanding of filing structure and organization.

GPS/PDR Appendix B. NPS Natural Resource Database Template Data Dictionary

The NRDT is a developing tool for data management that will change over time.

For the current version of the NRDT Data Dictionary refer to the NRDT web site (checked 9/27/2005): http://science.nature.nps.gov/im/apps/template/index.htm

For hardcopy distribution of the NPS Natural Resource Database Template Data Dictionary as an attachment to this GPS/PDR Standard Operating Procedure, print and attach the latest version from the online link above. The latest version is filed at this location on the GRYN server:

G:\Natural_Resource_databases\Database_template\DATA DICTIONARY.DOC (version 10/9/2002)